

# Payoff-Based Approach to Learning Generalized Nash Equilibria in Convex Games

Tatiana Tatarenko and Maryam Kamgarpour

**Abstract**—We consider multi-agent decision making, where each agent optimizes its cost function subject to local and global constraints. Agents' actions belong to a compact convex Euclidean space and agents' cost functions are convex. Under the assumption that the game mapping is pseudo-monotone, we propose a distributed payoff-based algorithm to learn the Nash equilibria. In the payoff-based approach, each agent uses only information about its current cost value and the constraint value with its associated dual multiplier, to compute its next action. We prove convergence of the proposed algorithm to a Nash equilibrium leveraging established results on stochastic processes. Furthermore, we derive a convergence rate for the algorithm under strong monotonicity assumption on the game mapping.

**Index Terms**—learning in games, multi-agent decision making, distributed algorithms

## I. INTRODUCTION

Decision making in multi-agent systems arises in engineering applications ranging from electricity market to telecommunication and transportation networks Arslan *et al.* (2007); Saad *et al.* (2012); Scutari *et al.* (2006). Game theory provides a powerful framework for analyzing and optimizing decisions in multi-agent systems. The notion of an equilibrium in a game characterizes desirable and stable solutions to multi-agent decision making problems. In this work, we design a distributed learning algorithm to converge to Nash equilibria for a class of non-cooperative games modeled by convex objective functions and coupling constraints.

There is a large body of work on computation of Nash equilibria. The approaches differ mainly by the particular structure of agents' cost functions as well as the information available to each agent. In a so-called *potential game*, a central optimization problem can be formulated whose minimizers coincide with a subset of the Nash equilibria of the game. One can then leverage distributed optimization algorithms to compute the minima of the potential function Li & Marden (2013); Salehisadaghiani & Pavel (2014) despite agents limited information of others' cost functions or action sets. Such distributed algorithms have also been designed for *aggregative games* Jensen (2010); Paccagnan *et al.* (2016b). In general, for implementation of the aforementioned distributed algorithms agents need to communicate with each other or with a central coordinator. Furthermore, the structure of the cost

function of each agent or its derivative is assumed to be known to that agent.

As an alternative to deterministic distributed optimization approaches, learning approaches start with the assumption that the functional form of each agent's objective function or its derivative may not be known to itself nor to other agents. They attempt to compute Nash equilibria by sampling agents' actions from a set of probability distributions. These probability distributions are updated based on the information available in the system. Most of the past work has focused on algorithms that require the ability to evaluate the cost function for arbitrary agents' actions. For example, Marden *et al.* (2009a); Perkins *et al.* (2015); Tatarenko (2014, 2016b) have dealt with learning procedures requiring the so-called oracle-based information, where each agent can calculate its current cost given any action from its action set.

There are many practical situations in which agents do not know functional form of the objectives. Such situations often arise in electricity markets (unknown price functions or constraints) Marden *et al.* (2013); Tellidou & Bakirtzis (2007), network routing (unknown traffic demands/constraints) Dutta *et al.* (2011); Marden *et al.* (2009b), and sensor coverage problems (density function on the mission space) Zhu & Martínez (2013), just to name a few examples. In such applications, each agent can only observe its obtained payoffs and be aware of its local actions. In this case, the information structure is referred to as *payoff-based*. A payoff-based learning algorithm in potential games is proposed in Marden & Shamma (2012) with the guarantee of stochastic stability of potential function minimizers. However, to implement this payoff-based algorithm agents need to have some memory. Other algorithms requiring only payoff-based information and memory are proposed in Goto *et al.* (2012) and Zhu & Martínez (2013). These learning procedures also guarantee stochastic stability of potential function minimizers. Moreover, by tuning a time-dependent parameter the learning procedures converge to a distribution over potential function minimizers in total variation. Learning based approaches are also proposed in Pradelski & Young (2012) for non-potential games, where stochastic convergence to the Nash equilibrium maximizing social welfare is guaranteed.

The aforementioned payoff-based procedures are applicable to games with finite action spaces. For games with uncountable action spaces, a payoff-based approach was proposed in Tatarenko (2016a) in a potential game setting. That work considered agents with no memory and agents' action sets that are the whole space  $\mathbb{R}$ . It was proven that the payoff-based algorithm converges to local minima of the

T. Tatarenko is with the Control Methods and Robotics Lab Technical University Darmstadt, Darmstadt, Germany 64283, M. Kamgarpour is with the Automatic Control Laboratory, ETH Zürich, Switzerland.

This work was gratefully supported by the German Research Foundation (DFG) within the GRK 1362 "Cooperative, Adaptive and Responsive Monitoring of Mixed Mode Environments" (www.gkmm.de) and by M. Kamgarpour's ERC Starting Grant CONENE.

potential function. This result was generalized to arbitrary (not necessarily potential) games with pseudo-monotone maps in Tatarenko & Kamgarpour (2017). Furthermore, the action space of each agent was generalized from  $\mathbb{R}$  to decoupled compact convex subsets of  $\mathbb{R}^m$ . However, in the presence of global coupling constraints on agents' actions, the previously presented payoff-based approach can no longer be applied.

Despite considerable progress in learning algorithms for games, to the best of our knowledge, the work on payoff-based learning has not considered coupling in agents' action spaces. In several realistic scenarios in which players share resources each player's feasible set depends on the other players' strategies. For example, in an electricity market, there are coupling constraints due to the underlying physical electricity network. Similar constraints exist in a transportation or telecommunication network and general deregulated economy problems Scutari *et al.* (2012). The Nash equilibria in a game with coupling constraints are referred to as *generalized Nash equilibria*. Ensuring uniqueness of these equilibria and computing them is a challenging problem and a lively research topic Facchinei & Kanzow (2007). We focus on a subset of generalized Nash equilibrium problems in which the coupling constraint is shared among agents. In this case, one can formulate a variational equilibrium problem to characterize a subset of the generalized Nash equilibria, referred to as variational equilibria. In addition to computational tractability, these equilibria present desired properties from practical viewpoint. In particular, since in this equilibrium, the dual multipliers associated to the joint constraint is equal across all players, there is a well-defined cost associated to constraint violation.

The seminal work Rosen (1965) considered the above class of problems and developed a continuous-time algorithm for convergence to variational equilibria (referred to as normalized equilibria in Rosen (1965)). Authors in Yin *et al.* (2011) consider variational equilibria in monotone games and propose a primal dual distributed algorithm. This algorithm requires agents being able to evaluate their cost function at any given action. Similarly, the work in Paccagnan *et al.* (2016a) considers computation of variational equilibria for quadratic games with linear coupling constraints. A distributed algorithm is proposed, given the functional form of agent's objective is known to itself. The work in Zhu & Frazzoli (2016) develops a primal dual algorithm for learning generalized Nash equilibria, which is robust to communication failures. In that work it is assumed that the agents are able to evaluate their cost function for arbitrary points in their action spaces. Thus, the gradient of the cost function can be approximated online.

Our contributions in this paper are as follows. First, we develop a payoff-based approach for computing variational Nash equilibria in a class of convex games with pseudo-monotone game maps and jointly convex coupling constraints. Hence, we generalize results that require monotonicity of the game map, or knowledge on the cost functions, constraints, or their gradients. Second, we prove almost sure convergence of the algorithm to variational Nash equilibria. Third, we quantify the convergence rate of the payoff-based algorithm, if the game map is strongly monotone.

Our approach is as follows. Similar to Paccagnan *et al.*

(2016a); Yin *et al.* (2011); Zhu & Frazzoli (2016), we extend the original game to define a player corresponding to a dual multiplier of the coupling constraints. Then, we develop a decentralized sampling based approach, in which the probability distributions from which agents sample their actions are Gaussian, inspired by the literature on learning automata Thathachar & Sastry (2003). Motivated by projection based algorithms, the mean of the distribution is updated iteratively by each agent based only on its current payoff and projected on the constraint set. This has an interpretation of a stochastic projected gradient algorithm Shapiro *et al.* (2014). The dual player, on the other hand, updates its action deterministically by measuring constraint violation at each time step. Notice that the dual player is a fictitious player. It can refer to a central coordinator who measures the constraint violation at each step. Alternatively, if each agent can locally measure constraint violation, then it can update its dual variable. To prove convergence of the proposed payoff-based algorithm we leverage results on Robbins-Monro stochastic approximation Bharath & Borkar (1999); Nevelson & Khasminskii (1973). Finally, the convergence rate is quantified based on rate estimates in stochastic projection algorithms Shapiro *et al.* (2014).

This paper is organized as follows. In Section II, we set up the game under consideration. In Section III, we propose our payoff-based approach and present its convergence result. Section IV develops the proof of the main result using supporting theorems on stochastic random variables. In Section V, we prove convergence rate of the algorithm under an additional assumption on the game setup. In Section VI, we summarize the result and discuss future work.

**Notations and basic definitions.** The set  $\{1, \dots, N\}$  is denoted by  $[N]$ . Boldface is used to distinguish between the vectors in a multi-dimensional space and scalars. Given  $N$  vectors  $\mathbf{x}^i \in \mathbb{R}^d$ ,  $i \in [N]$ ,  $[\mathbf{x}^i]_{i=1}^N := [\mathbf{x}^{1\top}, \dots, \mathbf{x}^{N\top}]^\top \in \mathbb{R}^{Nd}$ ,  $\mathbf{x}^{-i} := [\mathbf{x}^1, \dots, \mathbf{x}^{i-1}, \mathbf{x}^{i+1}, \dots, \mathbf{x}^N] \in \mathbb{R}^{(N-1)d}$ .  $\mathbb{R}_+^d$  and  $\mathbb{Z}_+$  denote the sets of vectors from  $\mathbb{R}^d$  with non-negative coordinates and non-negative whole numbers, respectively.  $\mathbb{R}_{\leq K}^d$  is the set of vectors from  $\mathbb{R}^d$  with norms that do not exceed  $K$ , namely  $\mathbb{R}_{\leq K}^d = \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\| \leq K\}$ . The standard inner product on  $\mathbb{R}^d$  is denoted by  $(\cdot, \cdot) : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ , with associated norm  $\|\mathbf{x}\| := \sqrt{(\mathbf{x}, \mathbf{x})}$ .  $I_d$  represents the  $d$ -dimensional identity matrix and  $\mathbf{1}_N$  represents the  $N$ -dimensional vector of unit entries. Given some matrix  $A \in \mathbb{R}^{d \times d}$ ,  $A \succeq (>)0$ , if and only if  $\mathbf{x}^\top A \mathbf{x} \geq (>)0$  for all  $\mathbf{x} \neq 0$ . Given a function  $\mathbf{g}(\mathbf{x}, \mathbf{y}) : \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \rightarrow \mathbb{R}$ , we define the mapping  $\nabla_{\mathbf{x}} \mathbf{g}(\mathbf{x}, \mathbf{y}) : \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \rightarrow \mathbb{R}^{d_1}$  component-wise as  $[\nabla_{\mathbf{x}} \mathbf{g}(\mathbf{x}, \mathbf{y})]_i := \frac{\partial \mathbf{g}(\mathbf{x}, \mathbf{y})}{\partial x^i}$ . We will use the big- $O$  notation. Namely, the function  $f(x) : \mathbb{R} \rightarrow \mathbb{R}$  is  $O(\mathbf{g}(x))$  as  $x \rightarrow a$ ,  $f(x) = O(g(x))$  as  $x \rightarrow a$ , if  $\lim_{x \rightarrow a} \frac{|f(x)|}{|g(x)|} \leq K$  for some positive constant  $K$ . We say that a function  $f(\mathbf{x})$  grows not faster than a function  $g(\mathbf{x})$  as  $\mathbf{x} \rightarrow \infty$ , if there exists a positive constant  $Q$ :  $f(\mathbf{x}) \leq g(\mathbf{x})$  for any  $\mathbf{x} : \|\mathbf{x}\| \geq Q$ .

**Definition 1:** The mapping  $\mathbf{M} : \mathbb{R}^d \rightarrow \mathbb{R}^d$  is called *pseudo-monotone* over  $X \subseteq \mathbb{R}^d$ , if  $(\mathbf{M}(\mathbf{y}), \mathbf{x} - \mathbf{y}) \geq 0$  implies  $(\mathbf{M}(\mathbf{x}), \mathbf{x} - \mathbf{y}) \geq 0$  for every  $\mathbf{x}, \mathbf{y} \in X$ .

**Definition 2:** The mapping  $\mathbf{M} : \mathbb{R}^d \rightarrow \mathbb{R}^d$  is called *strongly*

monotone over  $X \subseteq \mathbb{R}^d$  with constant  $\kappa > 0$ , if  $(\mathbf{M}(\mathbf{x}) - \mathbf{M}(\mathbf{y}), \mathbf{x} - \mathbf{y}) \geq \kappa \|\mathbf{x} - \mathbf{y}\|^2$  for any  $\mathbf{x}, \mathbf{y} \in X$ .

**Definition 3:** The mapping  $\mathbf{M}(\mathbf{x}_1, \dots, \mathbf{x}_{k+1}) : \mathbb{R}^{d(k+1)} \rightarrow \mathbb{R}^l$ , where  $\mathbf{x}_i \in \mathbb{R}^d$ ,  $i \in [k+1]$ , is called *Lipschitz on  $X \subseteq \mathbb{R}^{dk}$  with respect to coordinates  $\mathbf{x}_1, \dots, \mathbf{x}_k$*  with a function  $L(\mathbf{x}_{k+1}) > 0$  defined on  $Y \subseteq \mathbb{R}^d$ , if  $\|\mathbf{M}(\mathbf{x}, \mathbf{x}_{k+1}) - \mathbf{M}(\mathbf{y}, \mathbf{x}_{k+1})\| \leq L(\mathbf{x}_{k+1}) \|\mathbf{x} - \mathbf{y}\|$  for every  $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_k), \mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_k) \in X$ .

**Definition 4:** For a convex constraint set  $\mathcal{C}$ , we assume existence of convex functions such that  $\mathcal{C} = \{\mathbf{x} \in \mathbb{R}^d : f^i(\mathbf{x}) \leq 0, i \in [m]\}$ . The *Slater's constraint qualification condition* for  $\mathcal{C}$  consists in existence of a feasible point  $\mathbf{x}^* \in \mathcal{C}$ , with strict feasibility  $f^i(\mathbf{x}^*) < 0$ , for  $i$  corresponding to inequality constraints, and furthermore  $\mathbf{x}^* \in \text{relint}(D)$ ,  $D = \cap_{i=1}^m \text{dom}(f^i)$ . Here,  $\text{relint}(D) = \{\mathbf{x} \in D : \exists r > 0, B(\mathbf{x}, r) \cap \text{aff}(D) \subseteq D\}$  is the relative interior of  $D$ , with  $B(\mathbf{x}, r)$  denoting the ball centered at  $\mathbf{x}$  with radius  $r$  and  $\text{aff}(D)$  corresponds to the affine hull of  $D$  Boyd & Vandenberghe (2004). Moreover, there exists an  $\mathbf{x} \in \text{relint}(D)$ , such that absolute value of all constraints is bounded away from zero.

## II. PROBLEM FORMULATION

### A. Convex games with coupling constraints

We are focused here on a game  $\Gamma(N, \{A_i\}, \{J_i\}, C)$  with  $N$  players. We assume that the action of the  $i$ th player is locally constrained to  $\mathbf{a}^i \in A_i \subset \mathbb{R}^d$  and that the vector of joint actions<sup>1</sup>,  $\mathbf{a} = [\mathbf{a}^1 \dots \mathbf{a}^N] \in \mathbf{A} = A_1 \times \dots \times A_N$ , has to belong to a *global coupling constraint set  $C$* , namely

$$\mathbf{a} \in C = \{\mathbf{a} \in \mathbf{A} : \mathbf{g}(\mathbf{a}) \leq \mathbf{0}\}, \quad (1)$$

for some multivariate continuously differentiable function  $\mathbf{g} : \mathbb{R}^{Nd} \rightarrow \mathbb{R}^n$  capturing coupling inequality constraints with *convex* coordinates  $g_i(\mathbf{a}), i \in [n]$ . In this case, we say that the game  $\Gamma$  has coupled actions. Let  $\mathcal{Q} = \mathbf{A} \cap C$ ,  $\mathcal{Q}^i(\mathbf{a}^{-i}) = \{\mathbf{a}^i \in A_i : \mathbf{g}(\mathbf{a}^i, \mathbf{a}^{-i}) \leq \mathbf{0}\}$ . The cost functions  $J_i : \mathbb{R}^{Nd} \rightarrow \mathbb{R}$  indicate the cost  $J_i(\mathbf{a})$  the agent  $i$  has to pay, given any joint action  $\mathbf{a} \in \mathcal{Q}$ . We proceed with introducing the following definitions of game mappings.

**Definition 5:** The mapping  $\mathbf{M} : \mathbb{R}^{Nd} \rightarrow \mathbb{R}^{Nd}$ , referred to as the *game mapping* of a game  $\Gamma(N, \{A_i\}, \{J_i\}, C)$  with coupled actions ( $C \subset \mathbb{R}^{Nd}$ ) or uncoupled actions ( $C = \mathbb{R}^{Nd}$ ), is defined by

$$\begin{aligned} \mathbf{M}(\mathbf{a}) &= [M_{1,1}(\mathbf{a}), \dots, M_{1,d}(\mathbf{a}), \dots, M_{N,1}(\mathbf{a}), \dots, M_{N,d}(\mathbf{a})]^\top, \\ M_{i,k}(\mathbf{a}) &= \frac{\partial J_i(\mathbf{a})}{\partial a_k^i}, \quad \mathbf{a} \in \mathcal{Q} = \mathbf{A} \cap C, i \in [N], k \in [d]. \end{aligned} \quad (2)$$

**Definition 6:** The mapping  $\mathbf{M}^0 : \mathbb{R}^{Nd+n} \rightarrow \mathbb{R}^{Nd+n}$ , referred to as the *extended game mapping* of a game

$\Gamma(N, \{A_i\}, \{J_i\}, C)$  with coupled actions, is defined by

$$\begin{aligned} \mathbf{M}^0(\mathbf{a}, \boldsymbol{\lambda}) &= [M_1^0(\mathbf{a}, \boldsymbol{\lambda}), \dots, M_N^0(\mathbf{a}, \boldsymbol{\lambda}), -\mathbf{g}(\mathbf{a})]^\top, \\ M_i^0(\mathbf{a}, \boldsymbol{\lambda}) &= [M_{i,1}^0(\mathbf{a}, \boldsymbol{\lambda}), \dots, M_{i,d}^0(\mathbf{a}, \boldsymbol{\lambda})], \quad i \in [N], \\ M_{i,k}^0(\mathbf{a}, \boldsymbol{\lambda}) &= \frac{\partial (J_i(\mathbf{a}) + (\boldsymbol{\lambda}, \mathbf{g}(\mathbf{a})))}{\partial a_k^i}, \quad \mathbf{a} \in \mathcal{Q} = \mathbf{A} \cap C, \\ &\quad i \in [N], \quad k \in [d]. \end{aligned} \quad (3)$$

We make the following assumptions regarding the game  $\Gamma$ .

**Assumption 1:** The game under consideration is *convex*. Namely, for all  $i \in [N]$  the set  $A_i$  is convex and compact, the cost function  $J_i(\mathbf{a}^i, \mathbf{a}^{-i})$  is defined on  $\mathbb{R}^{Nd}$ , continuously differentiable in  $\mathbf{a}$  and convex in  $\mathbf{a}^i$  for fixed  $\mathbf{a}^{-i}$ .

**Assumption 2:** The coordinates  $M_i^0(\mathbf{a}, \boldsymbol{\lambda}) : \mathbb{R}^{Nd+1} \rightarrow \mathbb{R}^d$  of extended mapping  $\mathbf{M}^0(\mathbf{a}, \boldsymbol{\lambda}) : \mathbb{R}^{Nd+n} \rightarrow \mathbb{R}^{Nd+n}$  of a game  $\Gamma(N, \{A_i\}, \{J_i\}, C)$  with coupled actions are *Lipschitz on  $\mathbb{R}^{Nd}$  with respect to coordinates  $\mathbf{a}$*  with a linear function  $L_i(\boldsymbol{\lambda})$  (see Definition 3). The function  $\mathbf{g}(\mathbf{a})$  is Lipschitz on  $\mathbb{R}^{Nd}$ . Moreover, the extended game mapping  $\mathbf{M}^0(\mathbf{a}, \boldsymbol{\lambda})$  is *pseudo-monotone on  $\mathcal{Q} \times \mathbb{R}_+^n = (\mathbf{A} \cap C) \times \mathbb{R}_+^n$* .

**Remark 1:** Since the extended mapping  $\mathbf{M}^0(\mathbf{a}, \boldsymbol{\lambda})$  is affine in  $\boldsymbol{\lambda}$ , the Lipschitz condition for  $\mathbf{M}^0(\mathbf{a}, \boldsymbol{\lambda})$  in Assumption 2 above holds if the coordinates  $M_i(\mathbf{a}), i \in [N]$ , of game mapping  $\mathbf{M}(\mathbf{a})$  and the functions  $\frac{\partial g_j(\mathbf{a})}{\partial a_k^i}, j \in [n], k \in [N], i \in [d]$ , are Lipschitz with respect to their argument  $\mathbf{a} = (\mathbf{a}^1, \dots, \mathbf{a}^N)$  with some constants  $l_i, l_{j,i}^k$ .

**Remark 2:** In certain applications, for example, electricity markets Paccagnan *et al.* (2016b), agents' cost functions are represented by quadratic functions and coupling constraints are captured by linear functions. Obviously, in this case, the extended game mapping is affine, namely  $\mathbf{M}^0(\mathbf{a}, \boldsymbol{\lambda}) = M[\mathbf{a}, \boldsymbol{\lambda}]^\top + \mathbf{m}$ , where  $M$  is a matrix of the size  $(Nd+n) \times (Nd+n)$  and  $\mathbf{m}$  is a vector in  $\mathbb{R}^{Nd+n}$ . Due to Corollary 2 in Gowda (1990), the affine mapping above is pseudo-monotone if and only if the matrix  $M$  is positive semidefinite. This is in particular fulfilled, if the quadratic forms of the cost functions are positive definite or semidefinite (see Paccagnan *et al.* (2016b) and Tatarenko & Kamgarpour (2017) respectively).

**Assumption 3:** The sets  $A_i, i \in [N]$ ,  $\mathbf{A}$ , and  $\mathcal{Q}$  satisfy the Slater's constraint qualification (see Definition 4).

A *generalized Nash equilibrium (GNE)* in a game  $\Gamma$  with coupled actions represents a joint action from which no player has any incentive to unilaterally deviate.

**Definition 7:** A point  $\mathbf{a}^* \in \mathcal{Q}$  is called a *generalized Nash equilibrium (GNE)* if for any  $i \in [N]$  and  $\mathbf{a}^i \in \mathcal{Q}^i(\mathbf{a}^{-i})$

$$J_i(\mathbf{a}^{*i}, \mathbf{a}^{*-i}) \leq J_i(\mathbf{a}^i, \mathbf{a}^{*-i}).$$

In the case of  $C = \mathbb{R}^{Nd}$  we have  $\mathcal{Q}^i(\mathbf{a}^{-i}) = \{\mathbf{a}^i : \mathbf{a}^i \in A_i\}$  and any action  $\mathbf{a}^*$  for which the inequality above holds is a *Nash equilibrium (NE)*.

We focus on learning a generalized Nash equilibrium in a game with coupled actions. We are interested in designing a *payoff-based algorithm*, which converges to a generalized Nash equilibrium in any game for which Assumptions 1-3 hold. The payoff-based information structure implies that agent  $i, i \in [N]$ , does not know the functional form of  $J_i$ . It can only observe its values (its payoff) for an action it plays.

<sup>1</sup> All results below are applicable for games with different dimensions  $\{d_i\}$  of the action sets  $\{A_i\}$ .



### B. Generalized Nash equilibria and Variational Inequalities

In this subsection, we prove that the set of *GNE* is nonempty, given fulfillment of Assumptions 1-3 for the game  $\Gamma(N, \{A_i\}, \{J_i\}, C)$  with coupled actions. This result will be obtained through connecting generalized Nash equilibria and solutions of variational inequalities. Moreover, in the following subsection, we model an uncoupled action game associated with the game  $\Gamma$  and establish the relation between its Nash equilibria and GNE of the game  $\Gamma$ . Existence of such an uncoupled action game will allow us to present a payoff-based approach to learning GNE in the initial game  $\Gamma$ .

**Definition 8:** Consider a mapping  $T(\cdot): \mathbb{R}^d \rightarrow \mathbb{R}^d$  and a set  $Y \subseteq \mathbb{R}^d$ . The *solution set*  $SOL(Y, T)$  to the *variational inequality problem*  $VI(Y, T)$  is a set of vectors  $\mathbf{y}^* \in Y$  such that  $(T(\mathbf{y}^*), \mathbf{y} - \mathbf{y}^*) \geq 0$ , for all  $\mathbf{y} \in Y$ .

The following theorem is the well-known result on the existence of  $SOL(Y, T)$ , see Corollary 2.2.5 in Pang & Facchinei (2003).

**Theorem 1:** Given  $VI(Y, T)$ , suppose that the set  $Y$  is compact, convex and that the mapping  $T$  is continuous. Then,  $SOL(Y, T)$  is nonempty and compact.

Now, let us assume the set  $Y$  in the theorem above is expressed by the following inequality constraints:  $Y = \{\mathbf{y} \in \mathbb{R}^d : \mathbf{h}(\mathbf{y}) = [h_1(\mathbf{y}), \dots, h_m(\mathbf{y})]^T \leq \mathbf{0}\}$ , where  $h_i: \mathbb{R}^d \rightarrow \mathbb{R}$ ,  $i \in [m]$ , are some convex functions defined on  $\mathbb{R}^d$ . Then, given convexity of the functions  $h_i$  and Slater's condition for the set  $Y$ , conditions analogous to the Karush-Kuhn-Tucker ones can be formulated for  $SOL(Y, T)$  (see Proposition 1.3.4 in Pang & Facchinei (2003)). Namely, the following theorem takes place.

**Theorem 2:** Under conditions of Theorem 1,  $\mathbf{y}^* \in SOL(Y, T)$  if and only if there exists  $\boldsymbol{\nu} \in \mathbb{R}^m$  such that

$$\begin{aligned} \mathbf{0} &= T(\mathbf{y}^*) + \sum_{i=1}^m (\nu_i, \nabla h_i(\mathbf{y}^*)), \\ 0 &= (\boldsymbol{\nu}, \mathbf{h}(\mathbf{y}^*)), \quad \boldsymbol{\nu} \geq \mathbf{0}, \quad \mathbf{h}(\mathbf{y}^*) \leq \mathbf{0}. \end{aligned} \quad (4)$$

Thus, given  $VI(Y, T)$  above, we can associate a multiplier  $\boldsymbol{\nu}$  with any solution  $\mathbf{y}^*$  of  $VI(Y, T)$ . The relation between  $\mathbf{y}^*$  and  $\boldsymbol{\nu}$  is expressed by the KKT conditions (4). We further call the vector  $(\mathbf{y}^*, \boldsymbol{\nu})$  a *KKT tuple*.

Next, we formulate the result that establishes the connection between the set of *GNE* in a game  $\Gamma$  with coupled actions and solutions of a certain variational inequality (see Theorem 2.1 in Facchinei *et al.* (2007)).

**Theorem 3:** Suppose Assumption 1 holds for a game  $\Gamma(N, \{A_i\}, \{J_i\}, C)$  with coupled actions  $\mathcal{Q} = \mathbf{A} \cap C$ . Then, if  $\mathbf{a}^* \in SOL(\mathcal{Q}, \mathbf{M})$ , where  $\mathbf{M}$  is the game mapping in (2), then  $\mathbf{a}^*$  is a GNE in the game  $\Gamma$ .

Moreover, according to Theorem 1, Assumption 1 together with Assumption 2 guarantee non-emptiness of  $SOL(\mathcal{Q}, \mathbf{M})$ . Since any continuous extended game mapping implies a continuous game mapping (see Definitions 5 and 6), we get the following result.

**Corollary 1:** Let  $\Gamma(N, \{A_i\}, \{J_i\}, C)$  be a game with coupled actions for which Assumptions 1 and 2 hold. Then, there exists at least one generalized Nash equilibrium in  $\Gamma$

that belongs to the set  $SOL(\mathcal{Q}, \mathbf{M})$ , where  $\mathbf{M}$  is the game mapping (see Definition 2) and  $\mathcal{Q} = \mathbf{A} \cap C$ .

Any generalized Nash equilibrium that coincides with the solution of the corresponding *VI* as it is stated in the corollary above is called a *variational equilibrium*.

Note that Assumptions 1 and 2 do not imply uniqueness of the variational Nash equilibrium in  $\Gamma(N, \{A_i\}, \{J_i\}, C)$ . To guarantee uniqueness of the solution of  $VI(\mathcal{Q}, \mathbf{M})$ , one needs to consider a more restrictive assumption, for example, strong monotonicity of the game mapping  $\mathbf{M}$  (see, for example, Pang & Facchinei (2003))<sup>2</sup>. In our paper we do not restrict our attention to the case of a unique variational equilibrium, but deal with a broader class of games admitting multiple variational equilibria and investigate convergence only to one of them, which is dependent on the initial condition.

### C. Associated Games with Uncoupled Actions

In the previous subsection we discussed existence of GNE for the game  $\Gamma$ , where the agents' actions are coupled. The result in Corollary 1 allows reformulation of the problem of finding a generalized Nash equilibrium as the problem of solving  $VI(\mathcal{Q}, \mathbf{M})$ . To define a distributed payoff-based approach to the latter problem, we formulate yet another equivalent problem that concerns finding a Nash equilibrium in a new game  $\Gamma_a(\mathbf{A} \times \mathbb{R}_+^n)$ , with uncoupled actions. We refer to this game as the *associated game* of the initial game  $\Gamma$ :

$$\Gamma_a(\mathbf{A} \times \mathbb{R}_+^n) = \Gamma_a(N+1, \{J_i^0\}_{i \in [N+1]}, \{\{A_i\}_{i \in [N]}, \mathbb{R}_+^n\}), \quad (5)$$

with  $N+1$  players and *uncoupled actions*, where the first  $N$  players are called regular and the  $(N+1)$ th player is called dual. The action sets of the regular players coincide with the local action sets  $\{A_i\}$  of the players in the initial game  $\Gamma$ , whereas the action set of the dual player is the set  $\mathbb{R}_+^n$  of real  $n$ -dimensional vectors with non-negative coordinates. The cost functions of the players in  $\Gamma_a(\mathbf{A} \times \mathbb{R}_+^n)$  are defined as follows:

$$\begin{aligned} J_i^0(\mathbf{a}^i, \mathbf{a}^{-i}, \boldsymbol{\lambda}) &= J_i(\mathbf{a}^i, \mathbf{a}^{-i}) + (\boldsymbol{\lambda}, \mathbf{g}(\mathbf{a}^i, \mathbf{a}^{-i})), \quad i \in [N], \\ J_{N+1}^0(\mathbf{a}, \boldsymbol{\lambda}) &= -(\boldsymbol{\lambda}, \mathbf{g}(\mathbf{a})). \end{aligned} \quad (6)$$

The cost function of each regular player  $i \in [N]$  in the game  $\Gamma_a(\mathbf{A} \times \mathbb{R}_+^n)$  is composed by two terms: the original cost function from the game  $\Gamma$  plus an additional term that depends on the strategy  $\boldsymbol{\lambda}$  of the dual player and on the influence of the current joint action in the coupling constraint expressed by the function  $\mathbf{g}$ . As  $\boldsymbol{\lambda} \geq \mathbf{0}$ , the latter can be interpreted as a term penalizing violations of the global constraint by the given joint action. The cost of the dual player in turn ensures that the complementarity condition associated to the coupled constraint is met. The idea to consider such associated game is supported by the next lemma (Lemma 3 in Paccagnan *et al.* (2016a)).

**Lemma 1:** Suppose Assumptions 1 and 3 hold for the game  $\Gamma(N, \{A_i\}, \{J_i\}, C)$  with coupled actions, where  $C$  is defined as in (1),  $\mathcal{Q} = \mathbf{A} \cap C$ , and  $\mathbf{M}$  is the mapping of  $\Gamma$ . Then

<sup>2</sup>Note that uniqueness of a variational Nash equilibrium in  $\Gamma$  does not imply uniqueness of generalized Nash equilibrium in  $\Gamma$ .

$[\mathbf{a}^*, \boldsymbol{\lambda}^*]$  is a Nash equilibrium of the game  $\Gamma_a(\mathcal{A} \times \mathbb{R}_+^n)$  associated with the game  $\Gamma$  if and only if  $(\mathbf{a}^*, \boldsymbol{\nu})$  is a KKT tuple for the  $VI(\mathcal{Q}, \mathcal{M})$ , where  $\boldsymbol{\lambda}^*$  is the coordinate of the multiplier  $\boldsymbol{\nu}$  corresponding to the constraint  $\mathbf{g}(\mathbf{a}) \leq \mathbf{0}$ .

The approach in showing the above is through writing the KKT conditions for the game  $\Gamma_a(\mathcal{A} \times \mathbb{R}_+^n)$ . According to Corollary 1, there exists  $\mathbf{a}^* \in \text{SOL}(\mathcal{Q}, \mathcal{M})$ , if Assumptions 1 and 2 hold. In this case, Corollary 1 implies that  $\mathbf{a}^*$  is a GNE. Furthermore, Theorem 2 implies existence of  $\boldsymbol{\nu}$  such that  $(\mathbf{a}^*, \boldsymbol{\nu})$  is a KKT tuple for  $VI(\mathcal{Q}, \mathcal{M})$ . Let  $\boldsymbol{\lambda}^*$  be the coordinate of  $\boldsymbol{\nu}$  corresponding to the constraint  $\mathbf{g}(\mathbf{a}) \leq \mathbf{0}$ . Then, according to Lemma 1,  $[\mathbf{a}^*, \boldsymbol{\lambda}^*]$  is a Nash equilibrium of the game  $\Gamma_a(\mathcal{A} \times \mathbb{R}_+^n)$ .

From Lemma 5.1 in Zhu & Frazzoli (2016), the vector  $[\mathbf{a}^*, \boldsymbol{\lambda}^*]$  has a bounded norm, given fulfilment of Assumptions 1 and 3. Thus, taking into account above results and also Proposition 1.4.2 in Pang & Facchinei (2003), we obtain the following theorem.

**Theorem 4:** Let  $\Gamma(N, \{A_i\}, \{J_i\}, C)$  with coupled actions  $\mathcal{Q} = \mathcal{A} \cap C$  for which Assumptions 1-3 hold. Then the following is fulfilled:

- 1) there exists a Nash equilibrium  $[\mathbf{a}^*, \boldsymbol{\lambda}^*]$  in the associated game  $\Gamma_a(\mathcal{A} \times \mathbb{R}_+^n)$  with uncoupled actions,
- 2) some vector  $[\mathbf{a}^*, \boldsymbol{\lambda}^*] \in \mathcal{A} \times \mathbb{R}_+^n$  is a Nash equilibrium in  $\Gamma_a(\mathcal{A} \times \mathbb{R}_+^n)$ , if and only if  $[\mathbf{a}^*, \boldsymbol{\lambda}^*] \in \text{SOL}(\mathcal{A} \times \mathbb{R}_+^n, \mathcal{M}^0)$ , where  $\mathcal{M}^0$  is the mapping in Definition 6,
- 3) the vector norm of any Nash equilibrium  $[\mathbf{a}^*, \boldsymbol{\lambda}^*]$  in  $\Gamma_a(\mathcal{A} \times \mathbb{R}_+^n)$  is bounded, namely there exists a constant  $K > 0$  such that  $\|\boldsymbol{\lambda}^*\| \leq K$ , i.e.  $\boldsymbol{\lambda}^* \in \mathbb{R}_{\leq K}^n$ .

Finally, we introduce an *associated bounded game*  $\Gamma_{ab}(\mathcal{A} \times \mathbb{R}_{\leq K+r}^n)$  with respect to the initial game  $\Gamma$ . This game can be obtained from the associated game  $\Gamma_a(\mathcal{A} \times \mathbb{R}_+^n)$  defined by (5) and (6), where the action set  $\mathbb{R}_+^n$  of the  $(N+1)$ th dual player is replaced by the bounded set  $\mathbb{R}_{\leq K+r}^n$ , where  $K$  is defined in Theorem 4 and  $r$  is a finite positive constant. Note that Slater's condition in Assumption 3 allows us to use the result below that is proven in Proposition 3.1 and Lemma 5.1 in Zhu & Frazzoli (2016) for the game  $\Gamma_{ab}$ .

**Theorem 5:** Let  $\Gamma(N, \{A_i\}, \{J_i\}, C)$  be a game with coupled actions for which Assumptions 1 and 3 hold. Let  $\Gamma_{ab}(\mathcal{A} \times \mathbb{R}_{\leq K+r}^n)$  be its associated bounded game with uncoupled actions from the set  $\mathcal{A} \times \mathbb{R}_{\leq K+r}^n$ , where  $K$  is defined in Theorem 4 and  $r > 0$  is finite. Then, there exists a Nash equilibrium in  $\Gamma_{ab}(\mathcal{A} \times \mathbb{R}_{\leq K+r}^n)$ . Moreover, if  $[\mathbf{a}^*, \boldsymbol{\lambda}^*]$  is a Nash equilibrium in  $\Gamma_{ab}(\mathcal{A} \times \mathbb{R}_{\leq K+r}^n)$ , then  $\mathbf{a}^*$  is a GNE of  $\Gamma$ .

Moreover, Proposition 1.4.2 in Pang & Facchinei (2003) allows us to formulate the result for the game  $\Gamma_{ab}$  analogous to one for the game  $\Gamma_a$  in Theorem 4, assertion 2).

**Theorem 6:** Let  $\Gamma(N, \{A_i\}, \{J_i\}, C)$  be a game with coupled actions for which Assumption 1 holds. Let  $\Gamma_{ab}(\mathcal{A} \times \mathbb{R}_{\leq K+r}^n)$  be its associated bounded game with the uncoupled actions from the set  $\mathcal{A} \times \mathbb{R}_{\leq K+r}^n$ , where  $K$  is defined in Theorem 4 and  $r > 0$  is bounded. Then  $[\mathbf{a}^*, \boldsymbol{\lambda}^*]$  is a Nash equilibrium in  $\Gamma_{ab}(\mathcal{A} \times \mathbb{R}_{\leq K+r}^n)$  if and only if  $[\mathbf{a}^*, \boldsymbol{\lambda}^*] \in \text{SOL}(\mathcal{A} \times \mathbb{R}_{\leq K+r}^n, \mathcal{M}^0)$ .

The theorems above claim that to obtain a GNE in the initial game with coupled actions it suffices to find a Nash

equilibrium in the associated bounded game with uncoupled actions, which in its turn is equivalent to solving variational inequality  $VI(\mathcal{A} \times \mathbb{R}_{\leq K+r}^n, \mathcal{M}^0)$ . These equivalences enable design of distributed algorithms to converge to a desired equilibrium, given any initial agents' actions.

### III. SOLUTION APPROACH

In this section we consider the problem of payoff-based learning of a generalized Nash equilibrium in a game  $\Gamma(N, \{A_i\}, \{J_i\}, C)$  with coupled actions. Based on the discussion in the previous section, we focus on learning Nash equilibria in the associated bounded game  $\Gamma_{ab}(\mathcal{A} \times \mathbb{R}_{\leq K+r}^n)$  and its properties formulated in Theorem 5. To present an efficient payoff-based learning algorithm, we need to introduce the following assumption on the behavior of the cost functions at infinity.

**Assumption 4:** The cost functions  $J_i(\mathbf{a})$ ,  $i \in [N]$ , grow not faster than a quadratic function of  $\mathbf{a}$  as  $\|\mathbf{a}\| \rightarrow \infty$ .

**Remark 3:** Given Lipschitz continuity of  $\mathbf{g}$  on  $\mathbb{R}^{Nd}$  (see Assumption 2), the functions  $g_i(\mathbf{a})$ ,  $i \in [n]$ , grow not faster than a linear function of  $\mathbf{a}$  as  $\|\mathbf{a}\| \rightarrow \infty$ .

#### A. Payoff-Based Algorithm

In this subsection we formulate the payoff-based approach for the distributed learning of a Nash equilibrium  $\mathbf{a}^*$  in the associated game  $\Gamma_{ab}(\mathcal{A} \times \mathbb{R}_{\leq K+r}^n)$ , given fulfillment of Assumptions 1-4 for the initial game  $\Gamma$ .

Having access to information about the current state  $\mathbf{x}^i(t) = [x_1^i(t), \dots, x_d^i(t)]^\top \in \mathbb{R}^d$  at iteration  $t$  and the current cost value  $\hat{J}_i^0(t)$  at the joint state  $(\mathbf{x}(t), \boldsymbol{\lambda}(t))$ ,  $\hat{J}_i^0(t) = J_i^0(\mathbf{x}^1(t), \dots, \mathbf{x}^N(t), \boldsymbol{\lambda}(t))$ , each regular player  $i \in [N]$  "mixes" its next state  $\mathbf{x}^i(t+1)$ . Namely, it chooses its next state  $\mathbf{x}^i(t+1)$  randomly according to the multidimensional normal distribution  $\mathcal{N}(\boldsymbol{\mu}^i(t) = [\mu_1^i(t), \dots, \mu_d^i(t)]^\top, \sigma_i(t))$  with the density:

$$p_i(x_1^i, \dots, x_d^i; \boldsymbol{\mu}^i(t+1), \sigma_i(t+1)) = \frac{1}{(\sqrt{2\pi}\sigma_i(t+1))^d} \exp \left\{ - \sum_{k=1}^d \frac{(x_k^i - \mu_k^i(t+1))^2}{2\sigma_i^2(t+1)} \right\},$$

where  $i \in [N]$ . Our choice of Gaussian distribution is based on the idea of Continuous Action-set Learning Automaton (CALA), presented in the literature on learning automata Thathachar & Sastry (2003). The mean parameter  $\boldsymbol{\mu}^i(t)$  for the state's distribution is considered an action of the regular agent  $i$  at time step  $t$  and is updated as follows:

$$\boldsymbol{\mu}^i(t+1) = \text{Proj}_{A_i} \left[ \boldsymbol{\mu}^i(t) - \gamma_i(t+1)\sigma_i^2(t+1)\hat{J}_i^0(t) \frac{\mathbf{x}^i(t) - \boldsymbol{\mu}^i(t)}{\sigma_i^2(t)} \right] \quad (7)$$

where  $i \in [N]$  and  $\gamma_i(t+1)$  is a step size parameter chosen by player  $i$ ,  $\text{Proj}_{A_i}[\cdot]$  denotes the projection on set  $A_i$ . The initial value of  $\boldsymbol{\mu}(0)$  can be set to any finite value arbitrarily. We emphasize the difference between states and actions. In particular, the joint state at time  $t$  is an intermediary vector  $\mathbf{x}(t) = [\mathbf{x}^1(t), \dots, \mathbf{x}^N(t)]$  updated during the payoff-based algorithm under consideration. Unlike the actions  $\{\boldsymbol{\mu}^i(t)\}$ , the

states need not belong to the set of joint actions  $\mathbf{A}$  of the regular players. As will be shown, upon convergence of the algorithm, the states will also belong to the set of joint actions of the regular players.

As for the dual player  $N + 1$ , it updates its current action  $\lambda(t)$  based only on the observation of the violation of the constraint  $C$ , namely based on the actual value  $\hat{\mathbf{g}}(t)$ , of the function  $\mathbf{g}(\mathbf{x}^1(t), \dots, \mathbf{x}^N(t))$  at the current joint state of the regular players as follows:

$$\lambda(t + 1) = \text{Proj}_{\mathbb{R}_+^n}[\lambda(t) + \beta_0(t + 1)\hat{\mathbf{g}}(t)], \quad (8)$$

where  $\beta_0(t + 1)$  is a step size parameter chosen by the dual player  $N + 1$ . The initial value of  $\lambda(0)$  can be arbitrarily set to any finite value.

Note, in contrast to the approach in computing generalized Nash equilibria presented in Zhu & Frazzoli (2016), our proposed payoff-based algorithm does not rely on the specified bound  $K$  of the dual variable  $\lambda^*$  in the associated bounded game  $\Gamma_{ab}(\mathbf{A} \times \mathbb{R}_{\leq K+r}^n)$ . Indeed, notice that the update iterations for the dual variable  $\lambda(t)$  is projected onto the whole  $\mathbb{R}_+$  in (8). Nevertheless, the analysis below will demonstrate that the algorithm introduced above by (7)-(8) leads agents to a Nash equilibrium in the appropriate associated bounded game  $\Gamma_{ab}(\mathbf{A} \times \mathbb{R}_{\leq K+r}^n)$ .

### B. The Payoff Based Approach as a Stochastic Approximation Scheme

Our goal now is to analyze convergence of the proposed algorithm. First, we show that this algorithm is analogous to the Robbins-Monro stochastic approximation procedure Bharath & Borkar (1999). Next, we show the convergence of the random vector  $\eta(t) = [\mu(t), \lambda(t)] = [\mu^1(t), \dots, \mu^N(t), \lambda(t)]$  using existing results on stochastic processes and by properly choosing  $\{\beta_0(t)\}_{t=0}^\infty$  and  $\{\sigma_i(t), \gamma_i(t)\}_{t=0}^\infty$ ,  $i \in [N]$ .

Given  $\sigma = (\sigma_1, \dots, \sigma_N)$ , for any  $j \in [N + 1]$  define

$$\begin{aligned} \tilde{J}_j^0(\mu^1, \dots, \mu^N, \lambda, \sigma) &= \int_{\mathbb{R}^{Nd}} J_j^0(x, \lambda) p(\mu, x, \sigma) dx, \\ p(\mu, x, \sigma) &= \prod_{j=1}^N p_j(x_j^1, \dots, x_j^d; \mu^j, \sigma_j). \end{aligned}$$

Note that  $\tilde{J}_j^0$ ,  $j \in [N + 1]$ , can be interpreted as the  $j$ th player's cost function in mixed strategies of the regular players, given that the mixed strategies of these players are multivariate normal distributions  $\{\mathcal{N}(\mu^i, \sigma_i)\}_{i \in [N]}$ .

Under Assumptions 1 and 4 it follows that<sup>3</sup>

$$\begin{aligned} & \mathbb{E}_{\mathbf{x}(t)} \left\{ \hat{J}_i^0(t) \frac{x_k^i(t) - \mu_k^i(t)}{\sigma_i^2(t)} \right\} \\ &= \mathbb{E} \left\{ J_i^0(\mathbf{x}^1(t), \dots, \mathbf{x}^N(t), \lambda(t)) \frac{x_k^i(t) - \mu_k^i(t)}{\sigma_i^2(t)} \right\} \\ & \quad x_k^i(t) \sim \mathcal{N}(\mu_k^i(t), \sigma_i(t)), i \in [N], k \in [d] \\ &= \frac{\partial \tilde{J}_i^0(\mu^1(t), \dots, \mu^N(t), \lambda(t), \sigma(t))}{\partial \mu_k^i} \end{aligned}$$

<sup>3</sup>Assumptions 1 and 4 justify below differentiation under the integral sign of  $\tilde{J}_i^0(\mu^1(t), \dots, \mu^N(t), \lambda(t), \sigma(t))$ .

$$\text{for any } i \in [N], k \in [d], \quad (9)$$

$$\begin{aligned} \mathbb{E}_{\mathbf{x}(t)} \{\hat{\mathbf{g}}(t)\} &= \mathbb{E} \{\mathbf{g}(\mathbf{x}^1(t), \dots, \mathbf{x}^N(t))\} \\ x_k^i(t) &\sim \mathcal{N}(\mu_k^i(t), \sigma_i(t)), i \in [N], k \in [d] \\ &= -\nabla_{\lambda} \tilde{J}_{N+1}^0(\mu^1(t), \dots, \mu^N(t), \lambda(t), \sigma(t)). \end{aligned} \quad (10)$$

Using the notation  $\eta(t) = [\mu(t), \lambda(t)]$ , we can rewrite the algorithm steps (7)-(8) in the following form for  $i \in [N]$ :

$$\begin{aligned} \mu^i(t + 1) &= \text{Proj}_{\mathbf{A}}[\mu^i(t) - \gamma_i(t + 1)\sigma_i^2(t + 1) \\ &\quad \times (M_i^0(\eta(t)) + Q_i(\eta(t), \sigma(t)) \\ &\quad + R_i(\eta(t), \mathbf{x}(t), \sigma(t)))], \end{aligned} \quad (11)$$

$$\begin{aligned} \lambda(t + 1) &= \text{Proj}_{\mathbb{R}_+^n}[\lambda(t) - \beta_0(t + 1) \\ &\quad \times (-g(\mu(t)) + Q_{N+1}(\eta(t), \sigma(t)) \\ &\quad + R_{N+1}(\eta(t), \mathbf{x}(t), \sigma(t)))], \end{aligned} \quad (12)$$

where for  $i \in [N]$

$$\begin{aligned} Q_i(\eta(t), \sigma(t)) &= \tilde{M}_i^0(\eta(t), \sigma(t)) - M_i^0(\eta(t)), \\ R_i(\mathbf{x}(t), \eta(t), \sigma(t)) &= F_i(\mathbf{x}(t), \eta(t), \sigma(t)) - \tilde{M}_i^0(\eta(t), \sigma(t)), \\ F_i(\mathbf{x}(t), \eta(t), \sigma(t)) &= \hat{J}_i^0(t) \frac{\mathbf{x}^i(t) - \mu^i(t)}{\sigma_i^2(t)}, \end{aligned}$$

and

$$\tilde{M}_i^0(\cdot) = [\tilde{M}_{i,1}^0(\cdot), \dots, \tilde{M}_{i,d}^0(\cdot)]^\top \quad (13)$$

is the  $d$ -dimensional vector (mapping) with the following elements:

$$\tilde{M}_{i,k}^0(\eta(t), \sigma(t)) = \frac{\partial \tilde{J}_i^0(\eta(t), \sigma(t))}{\partial \mu_k^i}, \text{ for } k \in [d]. \quad (14)$$

Furthermore,

$$\begin{aligned} Q_{N+1}(\eta(t), \sigma(t)) &= \tilde{M}_{N+1}^0(\eta(t), \sigma(t)) + g(\mu(t)), \\ R_{N+1}(\mathbf{x}(t), \eta(t), \sigma(t)) &= -\hat{\mathbf{g}}(t) - \tilde{M}_{N+1}^0(\eta(t), \sigma(t)), \end{aligned}$$

and  $\tilde{M}_{N+1}^0(\eta(t), \sigma(t)) = \nabla_{\lambda} \tilde{J}_{N+1}^0(\eta(t), \sigma(t))$ .

It can then be observed that the algorithm (11)-(12) above falls under the framework of Robbins-Monro stochastic approximations procedure Bharath & Borkar (1999). Indeed, the vector

$$M^0(\eta(t)) = [M_1^0(\eta(t)), \dots, M_N^0(\eta(t)), -g(\mu(t))],$$

corresponds to the gradient term in stochastic approximation procedures, whereas

$$Q(\eta(t), \sigma(t)) = [Q_1(\eta(t), \sigma(t)), \dots, Q_N(\eta(t), \sigma(t)), Q_{N+1}(\eta(t), \sigma(t))],$$

is a disturbance of the gradient term, and

$$\begin{aligned} R(\mathbf{x}(t), \eta(t), \sigma(t)) &= [R_1(\mathbf{x}(t), \eta(t), \sigma(t)), \dots, \\ &\quad R_N(\mathbf{x}(t), \eta(t), \sigma(t)), R_{N+1}(\mathbf{x}(t), \eta(t), \sigma(t))] \end{aligned}$$



is a martingale difference. Namely, according to (9) and (10),

$$\begin{aligned} \mathbf{R}_i(\mathbf{x}(t), \boldsymbol{\eta}(t), \boldsymbol{\sigma}(t)) &= \mathbf{F}_i(\mathbf{x}(t), \boldsymbol{\eta}(t), \boldsymbol{\sigma}(t)) \\ &\quad - \mathbb{E}_{\mathbf{x}(t)}\{\mathbf{F}_i(\mathbf{x}(t), \boldsymbol{\eta}(t), \boldsymbol{\sigma}(t))\}, \end{aligned} \quad (15)$$

for  $i \in [N]$ , and

$$\mathbf{R}_{N+1}(\mathbf{x}(t), \boldsymbol{\eta}(t), \boldsymbol{\sigma}(t)) = -\hat{\mathbf{g}}(t) + \mathbb{E}_{\mathbf{x}(t)}\{\hat{\mathbf{g}}(t)\}. \quad (16)$$

Note that we assume the step size parameters to be individual for each agent. To be able to analyze the procedure we need to ensure the step sizes  $\beta_0(t)$ ,  $\sigma_i(t)$ ,  $\gamma_i(t)$ ,  $i \in [N]$ , satisfy certain assumptions. First, we introduce the following notations:  $\beta_i(t) = \gamma_i(t)\sigma_i^2(t)$ ,  $\beta_{\min}(t) = \min_{i \in \{0, [N]\}} \beta_i(t)$ ,  $\beta_{\max}(t) = \max_{i \in \{0, [N]\}} \beta_i(t)$ ,  $\gamma_{\max}(t) = \max_{i \in [N]} \gamma_i(t)$ .

*Assumption 5:* The variance parameters  $\sigma_i(t)$  and the step-size parameters  $\beta_0(t)$ ,  $\gamma_i(t)$ ,  $i \in [N]$ , are chosen such that

- 1)  $\sum_{t=0}^{\infty} \beta_{\min}(t) = \infty$ ,
- 2)  $\sum_{t=0}^{\infty} \beta_{\max}(t) - \beta_{\min}(t) < \infty$ ,
- 3)  $\sum_{t=0}^{\infty} \gamma_{\max}^2(t) < \infty$ ,  $\sum_{t=0}^{\infty} \gamma_{\max}(t)\sigma_{\max}^3(t) < \infty$ .

*Remark 4:* Note that condition 2) in Assumption 5 ensures some notion of consistency in the choice of individual players' parameters. An example of sequences  $\gamma_i(t)$ ,  $\sigma_i(t)$ ,  $i \in [N]$ ,  $\beta_0(t)$  can be the protocol for distributed optimization schemes, where each regular agent picks a positive integer  $R_i$ ,  $i \in [N]$ , the dual player picks a positive integer  $N_0$ , and the above conditions hold Kannan & Shanbhag (2012):  $\gamma_i(t) = \frac{1}{(t+R_i)^a}$ ,  $\sigma_i(t) = \frac{1}{(t+R_i)^b}$ ,  $i \in [N]$ ,  $\beta_0(t) = \frac{1}{(t+N_0)^{a+2b}}$ , where  $a+2b \in (0.5, 1]$ ,  $2a > 1$ , and  $a+3b > 1$  (for example,  $a = 0.6$ ,  $b = 0.2$ ).

Now, we are ready to formulate our main result on convergence of the payoff-based algorithm.

*Theorem 7:* Let Assumptions 1-5 hold in a game  $\Gamma(N, \{A_i\}, \{J_i\}, C)$  with coupled actions. Let the players in  $\Gamma$  be considered regular players in the associated bounded game  $\Gamma_{ab}(\mathcal{A} \times \mathbb{R}_{\leq K+r}^n)$  with uncoupled actions, who update their states  $\{\mathbf{x}^i(t)\}$  at time  $t$  according to the normal distribution  $\mathcal{N}(\boldsymbol{\mu}^i(t), \boldsymbol{\sigma}(t))$ , where the mean parameters correspond to the actions of the regular agents and are updated as in (7) and the action  $\boldsymbol{\lambda}(t)$  of the dual player is updated according to (8). Then, as  $t \rightarrow \infty$ , the mean vector  $\boldsymbol{\mu}(t)$  converges almost surely to a generalized Nash equilibrium  $\boldsymbol{\mu}^* = \mathbf{a}^*$  of the game  $\Gamma$ , given any initial vector  $[\boldsymbol{\mu}(0), \boldsymbol{\lambda}(0)]$ , and the joint state  $\mathbf{x}(t)$  converges in probability to  $\mathbf{a}^*$ .

According to the discussion in Subsection II-C, the theorem above claims almost sure convergence of the sequence of the mean vectors  $\{\boldsymbol{\mu}(t)\}$ , and, hence, agents' actions, and weak convergence of the sequence of the agents' states  $\{\mathbf{x}(t)\}$  to a generalized Nash equilibrium in the coupled action game under consideration. Note that, analogously to optimization methods based on the gradient descent iterations, condition 3) in Assumption 5  $\sum_{t=0}^{\infty} \gamma_{\max}(t)\sigma_{\max}^2(t) = \infty$ , guarantees sufficient energy for the time-step parameter  $\gamma_{\max}(t)\sigma_{\max}^2(t)$  to let the algorithm (11)-(12) get to a neighborhood of a desired stationary point, whereas condition  $\sum_{t=0}^{\infty} \gamma_{\max}^2(t) < \infty$  ensures the algorithm converges as time goes to infinity.

#### IV. ANALYSIS OF THE ALGORITHM CONVERGENCE

Our approach in proving Theorem (7) is to first prove boundedness of the iterates  $\boldsymbol{\eta}(t)$ . Next, we show that the limit

of the iterates  $\boldsymbol{\eta}(t)$  exists and satisfies the conditions for being a variational equilibrium of the game  $\Gamma(N, \{A_i\}, \{J_i\}, C)$ .

##### A. Supporting Theorems

In the following section, to prove convergence of the algorithm we will use the results on convergence properties of the Robbins-Monro stochastic approximation procedure analyzed in Nevelson & Khasminskii (1973).

We start by introducing some important notation. Let  $\{\mathbf{X}(t)\}_t$ ,  $t \in \mathbb{Z}_+$ , be a discrete-time Markov process on some state space  $E \subseteq \mathbb{R}^d$ , namely  $\mathbf{X}(t) = \mathbf{X}(t, \omega) : \mathbb{Z}_+ \times \Omega \rightarrow E$ , where  $\Omega$  is the sample space of the probability space on which the process  $\mathbf{X}(t)$  is defined. The transition function of this chain, namely  $\Pr\{\mathbf{X}(t+1) \in \Gamma | \mathbf{X}(t) = \mathbf{X}\}$ , is denoted by  $P(t, \mathbf{X}, t+1, \Gamma)$ ,  $\Gamma \subseteq E$ .

*Definition 9:* The operator  $L$  defined on the set of measurable functions  $V : \mathbb{Z}_+ \times E \rightarrow \mathbb{R}$ ,  $\mathbf{X} \in E$ , by

$$\begin{aligned} LV(t, \mathbf{X}) &= \int P(t, \mathbf{X}, t+1, dy)[V(t+1, y) - V(t, \mathbf{X})] \\ &= E[V(t+1, \mathbf{X}(t+1)) | \mathbf{X}(t) = \mathbf{X}] - V(t, \mathbf{X}), \end{aligned}$$

is called a *generating operator* of a Markov process  $\{\mathbf{X}(t)\}_t$ .

Now we formulate the following theorem for discrete-time Markov processes, which is proven in Nevelson & Khasminskii (1973), Theorem 2.5.2.

*Theorem 8:* Consider a Markov process  $\{\mathbf{X}(t)\}_t$  and suppose that there exists a function  $V(t, \mathbf{X}) \geq 0$  such that  $\inf_{t \geq 0} V(t, \mathbf{X}) \rightarrow \infty$  as  $\|\mathbf{X}\| \rightarrow \infty$  and

$$LV(t, \mathbf{X}) \leq -\alpha(t+1)\psi(t, \mathbf{X}) + f(t)(1 + V(t, \mathbf{X})),$$

where  $\psi \geq 0$  on  $\mathbb{R} \times \mathbb{R}^d$ ,  $f(t) > 0$ ,  $\sum_{t=0}^{\infty} f(t) < \infty$ . Let  $\alpha(t)$  be such that  $\alpha(t) > 0$ ,  $\sum_{t=0}^{\infty} \alpha(t) = \infty$ . Then, almost surely  $\sup_{t \geq 0} \|\mathbf{X}(t, \omega)\| = R(\omega) < \infty$ .

Furthermore, we will also need the next well-known result of Robbins and Siegmund on non-negative random variables, see, for example, Lemma 10 in Poljak (1987).

*Theorem 9:* Let  $(\Omega, F, P)$  be a probability space and  $F_1 \subset F_2 \subset \dots$  a sequence of sub- $\sigma$ -algebras of  $F$ . Let  $z_t, b_t, \xi_t$ , and  $\zeta_t$  be non-negative  $F_t$ -measurable random variables satisfying

$$\mathbb{E}(z_{t+1} | F_t) \leq z_t(1 + b_t) + \xi_t - \zeta_t.$$

Then, almost surely  $\lim_{t \rightarrow \infty} z_t$  exists and is finite. Moreover,  $\sum_{t=1}^{\infty} \zeta_t < \infty$  almost surely for the case in which  $\{\sum_{t=1}^{\infty} b_t < \infty, \sum_{t=1}^{\infty} \xi_t < \infty\}$ .

##### B. Boundedness of the Algorithm Iterates

First, we demonstrate that under conditions of Theorem 7 the vector  $\boldsymbol{\eta}(t)$  stays almost surely bounded for any  $t \in \mathbb{Z}_+$ .

*Lemma 2:* Let Assumptions 1-4 hold in a game  $\Gamma(N, \{A_i\}, \{J_i\}, C)$  with coupled actions and  $\boldsymbol{\eta}(t) = [\boldsymbol{\mu}(t), \boldsymbol{\lambda}(t)]$  be the vector updated in the run of the payoff-based algorithm (11)-(12). Let the variance parameters and the step-size parameters satisfy Assumption 5. Then,  $\Pr\{\sup_{t \geq 0} \|\boldsymbol{\eta}(t)\| < \infty\} = 1$ .

*Proof:* In the following, for simplicity in notation, we omit the argument  $\boldsymbol{\sigma}(t)$  in the terms  $M^0$ ,  $\tilde{M}^0$ ,  $Q$ , and  $R$ .

In certain derivations, for the same reason we omit the time parameter  $t$  as well. According to the vector form of the algorithm (11)-(12),

$$\begin{aligned} \mu^i(t+1) = & \text{Proj}_A[\mu^i(t) - \beta_i(t+1)(M_i^0(\eta(t)) \\ & + Q_i(\eta(t)) + R_i(\eta(t), \mathbf{x}(t)))], \end{aligned} \quad (17)$$

$$\begin{aligned} \lambda(t+1) = & \text{Proj}_{\mathbb{R}_+^n}[\lambda(t) - \beta_0(t+1) \\ & \times (-g(\mu(t)) + Q_{N+1}(\eta(t)) \\ & + R_{N+1}(\eta(t), \mathbf{x}(t)))], \end{aligned} \quad (18)$$

Let  $\eta^* = [\mu^*, \lambda^*] \in A \times \mathbb{R}_+^n$  be a Nash equilibrium of the associated game  $\Gamma_a(A \times \mathbb{R}_+^n)$ . This equilibrium exists and its norm is bounded, namely  $\|\eta^*\| < \infty$ , according to Theorem 4, assertions 1) and 3). Let us define the function  $V(\eta) = \|\eta - \eta^*\|^2$ . To prove the lemma we will consider the generating operator  $LV(\eta)$  and leverage the result in Theorem 8.

Taking into account the iterative procedure for the update of  $\eta(t)$  above and the non-expansion property of the projection operator on a convex set, we get

$$\begin{aligned} \|\mu_i(t+1) - \mu_i^*\|^2 &= \|\text{Proj}_A[\mu_i(t) - \beta_i(t+1)(M_i^0(\eta(t)) \\ & + Q_i(\eta(t)) + R_i(\mathbf{x}(t), \eta(t))) - \mu_i^*\|^2 \\ &\leq \|\mu_i(t) - \mu_i^* - \beta_i(t+1)(M_i^0(\eta(t)) \\ & + Q_i(\eta(t)) + R_i(\mathbf{x}(t), \eta(t)))\|^2 \\ &= \|\mu_i(t) - \mu_i^*\|^2 - 2\beta_i(t+1)(M_i^0(\eta(t)), \mu_i(t) - \mu_i^*) \\ & - 2\beta_i(t+1)(Q_i(\eta(t)) + R_i(\mathbf{x}(t), \eta(t)), \mu_i(t) - \mu_i^*) \\ & + \beta_i^2(t+1)\|G_i(\mathbf{x}(t), \eta(t))\|^2, \end{aligned} \quad (19)$$

where

$$G_i(\mathbf{x}(t), \eta(t)) = M_i^0(\eta(t)) + Q_i(\eta(t)) + R_i(\mathbf{x}(t), \eta(t)). \quad (20)$$

Similarly, we can bound the term corresponding to the dual player as follows

$$\begin{aligned} \|\lambda(t+1) - \lambda^*\|^2 &= \|\text{Proj}_{\mathbb{R}_+^n}[\lambda(t) - \beta_0(t+1)(-g(\mu(t)) + Q_{N+1}(\eta(t)) \\ & + R_{N+1}(\eta(t), \mathbf{x}(t))) - \lambda^*\|^2 \\ &\leq \|\lambda(t) - \lambda^* - \beta_0(t+1)(-g(\mu(t)) + Q_{N+1}(\eta(t)) \\ & + R_{N+1}(\mathbf{x}(t), \eta(t)))\|^2 \\ &= \|\lambda(t) - \lambda^*\|^2 - 2\beta_0(t+1)(-g(\mu(t)), \lambda(t) - \lambda^*) \\ & - 2\beta_0(t+1)(Q_{N+1}(\eta(t)) + R_{N+1}(\mathbf{x}(t), \eta(t)), \lambda(t) - \lambda^*) \\ & + \beta_0^2(t+1)\|G_{N+1}(\mathbf{x}(t), \eta(t))\|^2, \end{aligned} \quad (21)$$

with

$$\begin{aligned} G_{N+1}(\mathbf{x}(t), \eta(t)) &= -g(\mu(t)) + Q_{N+1}(\eta(t)) + R_{N+1}(\mathbf{x}(t), \eta(t)). \end{aligned} \quad (22)$$

Thus, taking into account the martingale properties (15) and (16) of the terms  $R_j$ ,  $j \in [N+1]$ , we obtain

$$\begin{aligned} LV(\eta) &= E[\|\eta(t+1) - \eta^*\|^2 | \eta(t) = \eta] - \|\eta - \eta^*\|^2 \\ &= \sum_{i=1}^N (E[\|\mu_i(t+1) - \mu_i^*\|^2 | \eta(t) = \eta] - \|\mu_i - \mu_i^*\|^2) \\ & \quad + E[\|\lambda(t+1) - \lambda^*\|^2 | \eta(t) = \eta] - \|\lambda - \lambda^*\|^2 \\ &\leq -2 \sum_{i=1}^N \beta_i(t+1)(M_i^0(\eta), \mu_i - \mu_i^*) \\ & \quad - 2\beta_0(t+1)(-g(\mu), \lambda - \lambda^*) \\ & \quad - 2 \sum_{i=1}^N \beta_i(t+1)(Q_i(\eta), \mu_i - \mu_i^*) \\ & \quad - 2\beta_0(t+1)(Q_{N+1}(\eta), \lambda - \lambda^*) \\ & \quad + \sum_{i=1}^N \beta_i^2(t+1)E\{\|G_i(\mathbf{x}(t), \eta(t))\|^2 | \eta(t) = \eta\} \\ & \quad + \beta_0^2(t+1)E\{\|G_{N+1}(\mathbf{x}(t), \eta(t))\|^2 | \eta(t) = \eta\}. \end{aligned} \quad (23)$$

Now, we bound the first two terms in the last expression above.

$$\begin{aligned} &-2 \sum_{i=1}^N \beta_i(t+1)(M_i^0(\eta), \mu_i - \mu_i^*) \\ & \quad - 2\beta_0(t+1)(-g(\mu), \lambda - \lambda^*) \\ &= -2\beta_{\min}(t+1)(M^0(\eta), \eta - \eta^*) \\ & \quad + 2\beta_{\min}(t+1)(M^0(\eta), \eta - \eta^*) \\ & \quad - 2 \sum_{i=1}^N \beta_i(t+1)(M_i^0(\eta), \mu_i - \mu_i^*) \\ & \quad - 2\beta_0(t+1)(-g(\mu), \lambda - \lambda^*) \\ &\leq -2\beta_{\min}(t+1)(M^0(\eta), \eta - \eta^*) \\ & \quad + 2(\beta_{\max}(t+1) - \beta_{\min}(t+1))\|M^0(\eta)\|\|\eta - \eta^*\| \\ &\leq -2\beta_{\min}(t+1)(M^0(\eta), \eta - \eta^*) \\ & \quad + 2(\beta_{\max}(t+1) - \beta_{\min}(t+1))k_1(1 + V(\eta)), \end{aligned} \quad (24)$$

for some constant  $k_1 > 0$ , where the last inequality is due to the linear behavior of the mapping  $M^0(\eta)$  at infinity (see Assumption 4). Hence, (23) and (24) imply

$$\begin{aligned} LV(\eta) &\leq -2\beta_{\min}(t+1)(M^0(\eta), \eta - \eta^*) \\ & \quad + 2(\beta_{\max}(t+1) - \beta_{\min}(t+1))k_1(1 + V(\eta)) \\ & \quad + 2 \sum_{i=1}^N \beta_i(t+1)\|Q_i(\eta)\|\|\mu_i - \mu_i^*\| \\ & \quad + 2\beta_0(t+1)\|Q_{N+1}(\eta)\|\|\lambda - \lambda^*\| \\ & \quad + \sum_{i=1}^N \beta_i^2(t+1)E\{\|G_i(\mathbf{x}(t), \eta(t))\|^2 | \eta(t) = \eta\} \\ & \quad + \beta_0^2(t+1)E\{\|G_{N+1}(\mathbf{x}(t), \eta(t))\|^2 | \eta(t) = \eta\}. \end{aligned} \quad (25)$$

Next, let us analyze the terms containing  $Q_i$  for  $i \in [N]$  and  $Q_{N+1}$  in (25). First, we will show that the mapping  $M_i^0(\eta(t))$



(see (13)-(14)) evaluated at  $\eta(t)$  is equivalent to the extended game mapping (see Definition 6) in mixed strategies, that is, for  $i \in [N+1]$

$$\tilde{M}_i^0(\eta(t)) = \int_{\mathbb{R}^{Nd}} M_i^0(x, \lambda) p(\mu(t), x) dx. \quad (26)$$

Using the notations

$$\mu_{-k}^i = (\mu_1^i, \dots, \mu_{k-1}^i, \mu_{k+1}^i, \dots, \mu_d^i) \in \mathbb{R}^{d-1},$$

$$x_{-k}^i = (x_1^i, \dots, x_{k-1}^i, x_{k+1}^i, \dots, x_d^i) \in \mathbb{R}^{d-1},$$

$$p(\mu_{-k}^i, x_{-k}^i) = \frac{1}{(\sqrt{2\pi}\sigma_i)^{d-1}} \exp \left\{ -\sum_{j \neq k} \frac{(x_j^i - \mu_j^i)^2}{2\sigma_i^2} \right\}$$

$$p(\mu^{-i}, x^{-i}) = \prod_{j \neq i, j=1}^N \frac{1}{(\sqrt{2\pi}\sigma_j)^d} \exp \left\{ -\sum_{k=1}^d \frac{(x_k^j - \mu_k^j)^2}{2\sigma_j^2} \right\},$$

we can show that for any  $i \in [N]$ ,  $k \in [d]$ ,  $\tilde{M}_{i,k}^0(\eta)$

$$\begin{aligned} \tilde{M}_{i,k}^0(\eta) &= \tilde{M}_{i,k}^0(\mu, \lambda) \\ &= \frac{1}{\sigma_i^2} \int_{\mathbb{R}^{Nd}} J_i^0(x, \lambda) (x_k^i - \mu_k^i) p(\mu, x) dx \\ &= - \int_{\mathbb{R}^{Nd}} J_i^0(x, \lambda) p(\mu_{-k}^i, x_{-k}^i) p(\mu^{-i}, x^{-i}) \frac{1}{\sqrt{2\pi}\sigma_i} \\ &\quad \times d \left( e^{-\frac{(x_k^i - \mu_k^i)^2}{2\sigma_i^2}} \right) dx^{-i} \\ &= - \int_{\mathbb{R}^{Nd-1}} \left( J_i^0(x, \lambda) e^{-\frac{(x_k^i - \mu_k^i)^2}{2\sigma_i^2}} \right) \Big|_{-\infty(x_k^i)}^{\infty(x_k^i)} \\ &\quad \times p(\mu_{-k}^i, x_{-k}^i) p(\mu^{-i}, x^{-i}) \frac{1}{\sqrt{2\pi}\sigma_i} dx^{-i} \\ &\quad + \int_{\mathbb{R}^{Nd}} \frac{\partial J_i^0(x, \lambda)}{\partial x_k^i} p(\mu, x) dx \\ &= \int_{\mathbb{R}^{Nd}} \frac{\partial J_i^0(x, \lambda)}{\partial x_k^i} p(\mu, x) dx, \end{aligned} \quad (27)$$

The above holds since, according to Assumption 4,

$$\lim_{x_k^i \rightarrow \infty(-\infty)} J_i^0(x, \lambda) e^{-\frac{(x_k^i - \mu_k^i)^2}{2\sigma_i^2}} = 0$$

for any fixed  $\mu_k^i$ ,  $x^{-i}$ , and  $\lambda$ . Thus, (26) holds for each regular player  $i \in [N]$ . Moreover, for the dual player

$$\begin{aligned} \tilde{M}_{N+1}^0(\eta) &= \int_{\mathbb{R}^{Nd}} \nabla_{\lambda} J_{N+1}^0(x, \lambda) p(\mu, x) dx \\ &= - \int_{\mathbb{R}^{Nd}} \mathbf{g}(x) p(\mu, x) dx. \end{aligned} \quad (28)$$

Since  $Q_i(\eta(t)) = \tilde{M}_i^0(\eta(t)) - M_i^0(\eta(t))$  and due to Assumption 2 and equation (26), we obtain the following:

$$\begin{aligned} \|Q_i(\eta)\| &= \left\| \int_{\mathbb{R}^{Nd}} [M_i^0(x, \lambda) - M_i^0(\mu, \lambda)] p(\mu, x) dx \right\| \\ &\leq \int_{\mathbb{R}^{Nd}} \|M_i^0(x, \lambda) - M_i^0(\mu, \lambda)\| p(\mu, x) dx \\ &\leq \int_{\mathbb{R}^{Nd}} L_i(\lambda) \|x - \mu\| p(\mu, x) dx \end{aligned}$$

$$\begin{aligned} &\leq \int_{\mathbb{R}^{Nd}} L_i(\lambda) \left( \sum_{i=1}^N \sum_{k=1}^d |x_k^i - \mu_k^i| \right) p(\mu, x) dx \\ &= O(\sigma_i)(1 + \|\eta - \eta^*\|), \end{aligned} \quad (29)$$

where the last equality is due to the fact that the first central absolute moment of a random variable with a normal distribution  $\mathcal{N}(\mu, \sigma)$  is  $O(\sigma)$  and  $L_i(\lambda)$  is a linear function of  $\lambda$  (see Assumption 2) and, hence,  $L_i(\lambda) \leq k(1 + \|\eta - \eta^*\|)$  for some constant  $k$ . Thus,

$$\|Q_i(\eta)\| \|\mu_i - \mu_i^*\| \leq O(\sigma_i)(1 + V(\eta)). \quad (30)$$

Analogously, using Assumption 2 and equality (28), we can show that

$$\begin{aligned} \|Q_{N+1}(\eta)\| &= \|\tilde{M}_{N+1}^0(\eta) + \mathbf{g}(\mu)\| \leq O\left(\sum_{i=1}^N \sigma_i\right), \\ \|Q_{N+1}(\eta)\| \|\lambda - \lambda^*\| &\leq O\left(\sum_{i=1}^N \sigma_i\right)(1 + V(\eta)). \end{aligned} \quad (31)$$

Finally, we bound the last two terms in (25). Since  $E(\xi - E\xi)^2 \leq E\xi^2$  and taking into account (15), we have

$$\begin{aligned} &E\{\|\mathbf{R}_i(\mathbf{x}(t), \eta(t))\|^2 | \eta(t) = \eta\} \\ &\leq \sum_{k=1}^d \int_{\mathbb{R}^{Nd}} J_i^{02}(x, \lambda) \frac{(x_k^i - \mu_k^i(t))^2}{\sigma_i^4(t)} p(\mu, x) dx. \end{aligned} \quad (32)$$

Thus, we can use Assumption 4, Remark 3, and the fact that  $J_i^0(x, \lambda)$  is affine in  $\lambda$  to get the next inequality:

$$\begin{aligned} &E\{\|\mathbf{R}_i(\mathbf{x}(t), \eta(t))\|^2 | \eta(t) = \eta\} \\ &\leq f(\mu, \sigma(t)) \left( \frac{1}{\sigma_i^4(t)} + k_2 V(\eta) \right), \end{aligned} \quad (33)$$

where  $f(\mu, \sigma(t))$  is a polynomial of  $\mu_i$  and  $\sigma_i(t)$ ,  $i \in [N]$ , and  $k_2$  is some positive constant.

Furthermore, taking into account boundedness of  $\mu(t)$  and affinity of the mapping  $M^0(\eta)$  with respect to  $\lambda$ , we obtain the following bound for any  $i \in [N]$ :

$$\begin{aligned} &\beta_i^2(t+1) E\{\|\mathbf{G}_i(\mathbf{x}(t), \eta(t))\|^2 | \eta(t) = \eta\} \\ &\leq \beta_i^2(t+1) (\|M_i^0(\eta)\|^2 + \|Q_i(\eta(t))\|^2) \\ &\quad + 2\beta_i^2(t+1) \|M_i^0(\eta)\| \|Q_i(\eta)\| \\ &\quad + \beta_i^2(t+1) (E\{\|\mathbf{R}_i(\mathbf{x}(t), \eta(t))\|^2 | \eta(t) = \eta\}) \\ &\leq \beta_i^2(t+1) (k_3 + k_4 \|\lambda\|^2 + O(\sigma_i^2(t))(1 + \|\eta - \eta^*\|)^2) \\ &\quad + 2\beta_i^2(t+1) (k_5 + k_6 \|\lambda\|) O(\sigma_i(t)(1 + \|\eta - \eta^*\|)) \\ &\quad + O(\gamma_i^2(t+1)) + O(\beta_i^2(t+1)) V(\eta) \\ &\leq O(\beta_i^2(t+1)(1 + \sigma_i^2(t) + \sigma_i(t) + \gamma_i^2(t+1))) \\ &\quad \times (1 + V(\eta)), \end{aligned} \quad (34)$$

where  $k_j$ ,  $j = 3, \dots, 6$ , are some positive constants.

For the term  $E\{\|\mathbf{G}_{N+1}(\mathbf{x}(t), \eta(t))\|^2 | \eta(t) = \eta\}$  we can first derive the following bound:

$$\begin{aligned} &E\{\|\mathbf{R}_{N+1}(\mathbf{x}(t), \eta(t))\|^2 | \eta(t) = \eta\} \\ &\leq \int_{\mathbb{R}^{Nd}} \|\mathbf{g}(x)\|^2 p(\mu, x) dx = \phi(\mu, \sigma(t)), \end{aligned}$$

where  $\phi(\boldsymbol{\mu}, \boldsymbol{\sigma}(t))$  is a polynomial of  $\mu_i$  and  $\sigma_i(t)$ ,  $i \in [N]$  (see Remark 3). Hence, according to boundedness of  $\boldsymbol{\mu}(t)$  and the fact that  $\sigma(t)$  goes to zero, we obtain

$$\begin{aligned} & \beta_0^2(t+1)\mathbb{E}\{\|\mathbf{G}_{N+1}(\boldsymbol{\eta}(t))\|^2|\boldsymbol{\eta}(t) = \boldsymbol{\eta}\} \\ & \leq \beta_0^2(t+1)(\|\mathbf{g}(\boldsymbol{\mu})\|^2 + \|\mathbf{Q}_{N+1}(\boldsymbol{\eta}(t))\|^2) \\ & \quad + 2\beta_0^2(t+1)\|\mathbf{g}(\boldsymbol{\mu})\|\|\mathbf{Q}_{N+1}(\boldsymbol{\eta})\| \\ & \quad + \beta_0^2(t+1)(\mathbb{E}\{\|\mathbf{R}_{N+1}(\boldsymbol{\eta}(t))\|^2|\boldsymbol{\eta}(t) = \boldsymbol{\eta}\}) \\ & \leq \beta_0^2(t+1)(k_7 + O(\sigma_{\max}^2(t)) + O(\sigma_{\max}(t))), \end{aligned} \quad (35)$$

where  $k_7$  is some positive constant.

Since  $\boldsymbol{\eta}^*$  is a NE in  $\Gamma_a(\mathcal{A} \times \mathbb{R}_+^n)$ , assertion 2) in Theorem 4 implies that

$$(M^0(\boldsymbol{\eta}^*), \boldsymbol{\eta} - \boldsymbol{\eta}^*) \geq 0$$

for any  $\boldsymbol{\eta} \in \mathcal{A} \times \mathbb{R}_+^n$ . According to pseudo-monotonicity of  $M^0$  in Assumption 2, the inequality above implies

$$(M^0(\boldsymbol{\eta}), \boldsymbol{\eta} - \boldsymbol{\eta}^*) \geq 0 \text{ for any } \boldsymbol{\eta} \in \mathcal{A} \times \mathbb{R}_+^n. \quad (36)$$

Thus, bringing (25), (30), (31), (34), and (35) together, we can write

$$\begin{aligned} LV(\boldsymbol{\eta}) & \leq -2\beta_{\min}(t+1)(M^0(\boldsymbol{\eta}), \boldsymbol{\eta} - \boldsymbol{\eta}^*) \\ & \quad + p(t)(1 + V(\boldsymbol{\eta})), \end{aligned} \quad (37)$$

where

$$\begin{aligned} p(t) & = O(\beta_{\max}(t+1) - \beta_{\min}(t+1)) \\ & \quad + O(\beta_{\max}(t+1)\sigma_{\max}(t) + \gamma_{\max}^2(t+1)). \end{aligned}$$

Hence, using conditions on parameters in Assumption 5, we get  $\sum_{t=0}^{\infty} p(t) < \infty$ . Finally, taking into account (36), (37),  $\sum_{t=0}^{\infty} \beta_{\min}(t) = \infty$ , and Theorem 8, we conclude that  $\boldsymbol{\eta}(t)$  is finite almost surely for any  $t \in \mathbb{Z}_+$  during the run of the algorithm irrespective of  $\boldsymbol{\eta}(0)$ . ■

### C. Convergence of the Algorithm Iterates

To prove Theorem 7 we can now use a reasoning analogous to the one in the proof of Theorem 2 in Tatarenko & Kamgarpour (2017). Note that, according to Theorem 5, a sufficient condition for a joint action  $\mathbf{a}^*$  to be a GNE in a game with coupled actions is that the vector  $[\mathbf{a}^*, \boldsymbol{\lambda}^*]$  is a Nash equilibrium in the associated bounded game  $\Gamma_{ab}(\mathcal{A} \times \mathbb{R}_{\leq K+r}^n)$  (see Theorem 5).

*Proof of Theorem 7:* Our approach is as follows. First, we estimate the distance between the vector  $\boldsymbol{\eta}(t+1) = [\boldsymbol{\mu}(t+1), \boldsymbol{\lambda}(t+1)]$  in the run of the algorithm and some arbitrary  $\boldsymbol{\eta} = [\boldsymbol{\mu}, \boldsymbol{\lambda}] \in \mathcal{A} \times \mathbb{R}_{\leq K+r}^n$ , where  $K$  is defined in Theorem 4 and  $r$  is a positive bounded constant specified below, by the same distance in the previous step, namely by  $\|\boldsymbol{\eta}(t) - \boldsymbol{\eta}\|$ . After that, we analyse each term in this estimation to demonstrate applicability of Theorem 9 to the sequence  $\{\boldsymbol{\eta}(t)\}_t$ . Finally, we use the properties of Nash equilibria in games satisfying Assumptions 1-5 (see Theorem 4) to demonstrate that the almost sure limit of the sequence  $\{\boldsymbol{\eta}(t)\}_t$  is a Nash equilibrium in the uncoupled action game  $\Gamma_{ab}(\mathcal{A} \times \mathbb{R}_{\leq K+r}^n)$ , that is, the associated bounded game with respect to the coupled action game  $\Gamma$  under consideration.

According to Lemma 2,  $\boldsymbol{\eta}(t)$  is bounded for any  $t$ . Hence, there exists a point  $\boldsymbol{\eta}^* = [\boldsymbol{\mu}^*, \boldsymbol{\lambda}^*] \in \mathcal{A} \times \mathbb{R}_+^n$  with a bounded norm such that  $\lim_{t \rightarrow \infty} \boldsymbol{\eta}(t) = \boldsymbol{\eta}^*$ . Furthermore, we specify the bounded constant  $r > 0$  in the game  $\Gamma_{ab}(\mathcal{A} \times \mathbb{R}_{\leq K+r}^n)$  to fulfill the following condition:

$$\|\boldsymbol{\lambda}^*\| \leq K + r. \quad (38)$$

Note that due to the boundedness of  $\boldsymbol{\eta}^* = [\boldsymbol{\mu}^*, \boldsymbol{\lambda}^*]$  such  $r$  always exists and is finite. Let  $\boldsymbol{\eta} = [\boldsymbol{\mu}, \boldsymbol{\lambda}] \in \mathcal{A} \times \mathbb{R}_{\leq K+r}^n$  be any bounded vector from the joint action set of the game  $\Gamma_{ab}(\mathcal{A} \times \mathbb{R}_{\leq K+r}^n)$ . Then, analogous to (19) and (39), we get

$$\begin{aligned} & \|\boldsymbol{\eta}(t+1) - \boldsymbol{\eta}\|^2 \\ & \leq \|\boldsymbol{\eta}(t) - \boldsymbol{\eta}\|^2 - 2 \sum_{i=1}^N \beta_i(t+1)(M^0(\boldsymbol{\eta}(t)), \boldsymbol{\mu}_i(t) - \boldsymbol{\mu}_i) \\ & \quad - 2\beta_0(t+1)(-\mathbf{g}(\boldsymbol{\mu}(t)), \boldsymbol{\lambda}(t) - \boldsymbol{\lambda}) \\ & \quad - 2 \sum_{i=1}^N \beta_i(t+1)(\mathbf{Q}_i(\boldsymbol{\eta}) + \mathbf{R}_i(\mathbf{x}(t), \boldsymbol{\eta}(t)), \boldsymbol{\mu}_i(t) - \boldsymbol{\mu}_i) \\ & \quad - 2\beta_0(t+1)(\mathbf{Q}_{N+1}(\boldsymbol{\eta}) + \mathbf{R}_{N+1}(\mathbf{x}(t), \boldsymbol{\eta}(t)), \boldsymbol{\lambda}(t) - \boldsymbol{\lambda}^*) \\ & \quad + \sum_{i=1}^N \beta_i^2(t+1)\|\mathbf{G}_i(\mathbf{x}(t), \boldsymbol{\eta}(t))\|^2 \\ & \quad + \beta_0^2(t+1)\|\mathbf{G}_{N+1}(\mathbf{x}(t), \boldsymbol{\eta}(t))\|^2. \end{aligned} \quad (39)$$

Let  $\mathcal{F}_T$  be the  $\sigma$ -algebra generated by the random variables  $\{\boldsymbol{\eta}(k), k \leq T\}$ . By taking the conditional expectation with respect to  $\mathcal{F}_T$  of the both sides in the inequality above, we obtain that for any  $T > 0$ , almost surely

$$\begin{aligned} & 2 \sum_{t=0}^T \sum_{i=1}^N \beta_i(t+1)(M_i^0(\boldsymbol{\eta}(t)), \boldsymbol{\mu}_i(t) - \boldsymbol{\mu}_i) \\ & \quad + 2 \sum_{t=0}^T \beta_0(t+1)(-\mathbf{g}(\boldsymbol{\mu}(t)), \boldsymbol{\lambda}(t) - \boldsymbol{\lambda}) \\ & \leq \|\boldsymbol{\eta}(0) - \boldsymbol{\eta}\|^2 - \mathbb{E}\{\|\boldsymbol{\eta}(T+1) - \boldsymbol{\eta}\|^2 | \mathcal{F}_T\} \\ & \quad + 2 \sum_{t=0}^T \sum_{i=1}^N \beta_i(t+1)\|\mathbf{Q}_i(\boldsymbol{\eta}(t))\|\|\boldsymbol{\mu}_i(t) - \boldsymbol{\mu}_i\| \\ & \quad + 2 \sum_{t=0}^T \beta_0(t+1)\|\mathbf{Q}_{N+1}(\boldsymbol{\eta}(t))\|\|\boldsymbol{\lambda}(t) - \boldsymbol{\lambda}\| \\ & \quad + \sum_{t=0}^T \sum_{i=1}^N \beta_i^2(t+1)\mathbb{E}_{\mathbf{x}(t)}\{\|\mathbf{G}_i(\mathbf{x}(t), \boldsymbol{\eta}(t))\|^2\} \\ & \quad + \sum_{t=0}^T \beta_0^2(t+1)\mathbb{E}_{\mathbf{x}(t)}\{\|\mathbf{G}_{N+1}(\mathbf{x}(t), \boldsymbol{\eta}(t))\|^2\}, \end{aligned} \quad (40)$$

where for  $j \in [N+1]$

$$\begin{aligned} & \mathbb{E}_{\mathbf{x}(t)}\{\|\mathbf{G}_j(\mathbf{x}(t), \boldsymbol{\eta}(t))\|^2\} \\ & = \mathbb{E}\{\|\mathbf{G}_j(\mathbf{x}(t), \boldsymbol{\eta}(t))\|^2 | x_k^i(t) \sim \mathcal{N}(\mu_k^i(t), \sigma_i(t)), \\ & \quad i \in [N], k \in [d]\}. \end{aligned}$$

In inequality (40) we used the Cauchy-Schwarz inequality as well as the property of the conditional expectation, namely  $\mathbb{E}\{\boldsymbol{\eta}(t_1) | \mathcal{F}_{t_2}\} = \boldsymbol{\eta}(t_1)$  almost surely for any  $t_1 \leq t_2$ , the fact that  $\mathbb{E}\{\mathbf{R}_j(\mathbf{x}(t), \boldsymbol{\eta}(t)) | \mathcal{F}_T\} = 0$  for all  $j \in [N+1]$ ,  $t \leq T$ ,

which is implied by (15), (16). Analogously to (30) and (31), we obtain

$$\begin{aligned} \|\mathbf{Q}_i(\boldsymbol{\eta}(t))\| &\leq O(\sigma_i(t)), \\ \|\mathbf{Q}_i(\boldsymbol{\eta}(t))\| \|\boldsymbol{\mu}(t) - \boldsymbol{\mu}\| &\leq O(\sigma_i(t)). \end{aligned} \quad (41)$$

$$\begin{aligned} \|\mathbf{Q}_{N+1}(\boldsymbol{\eta}(t))\| &\leq O(\sigma_{\max}(t)), \\ \|\mathbf{Q}_{N+1}(\boldsymbol{\eta}(t))\| \|\boldsymbol{\lambda}(t) - \boldsymbol{\lambda}\| &\leq O(\sigma_{\max}(t)). \end{aligned} \quad (42)$$

The inequalities (41) and (42) above are due to boundedness of  $\|\boldsymbol{\eta}(t) - \boldsymbol{\eta}\|$  implied by Lemma 2. Now we proceed with estimating the terms  $\mathbb{E}_{\mathbf{x}(t)} \|\mathbf{G}_j(\mathbf{x}(t), \boldsymbol{\eta}(t))\|^2$ ,  $j \in [N+1]$  in (40). According to the definition of  $\mathbf{G}_j$  in (20), (22), and from the Cauchy-Schwarz inequality,

$$\begin{aligned} \mathbb{E}_{\mathbf{x}(t)} \|\mathbf{G}_i(\mathbf{x}(t), \boldsymbol{\eta}(t))\|^2 &\leq \|\mathbf{M}_i^0(\boldsymbol{\eta}(t))\|^2 + \|\mathbf{Q}_i(\boldsymbol{\eta}(t))\|^2 \\ &\quad + 2\|\mathbf{M}_i^0(\boldsymbol{\eta}(t))\| \|\mathbf{Q}_i(\boldsymbol{\eta}(t))\| \\ &\quad + \mathbb{E}\{\|\mathbf{R}_i(\mathbf{x}(t), \boldsymbol{\eta}(t))\|^2\}, \quad i \in [N], \end{aligned} \quad (43)$$

$$\begin{aligned} \mathbb{E}_{\mathbf{x}(t)} \|\mathbf{G}_{N+1}(\mathbf{x}(t), \boldsymbol{\eta}(t))\|^2 &\leq \|\mathbf{g}(\boldsymbol{\mu}(t))\|^2 + \|\mathbf{Q}_{N+1}(\boldsymbol{\eta}(t))\|^2 \\ &\quad + 2\|\mathbf{g}(\boldsymbol{\mu}(t))\| \|\mathbf{Q}_{N+1}(\boldsymbol{\eta}(t))\| \\ &\quad + \mathbb{E}\{\|\mathbf{R}_{N+1}(\mathbf{x}(t), \boldsymbol{\eta}(t))\|^2\}. \end{aligned} \quad (44)$$

Furthermore, taking into account boundedness of  $\boldsymbol{\eta}(t)$  for all  $t$ , Assumption 2, and (33), we conclude that for any  $i \in [N]$

$$\beta_i^2(t+1) \mathbb{E}\{\|\mathbf{R}_i(\mathbf{x}(t), \boldsymbol{\eta}(t))\|^2\} \leq O(\sigma_{\max}^2(t) \gamma_{\max}^2(t)). \quad (45)$$

Using again boundedness of  $\boldsymbol{\eta}(t)$  for all  $t$ , we conclude that the first three terms on the right hand side of (43) and (44) are bounded.

Bringing (35), (41) - (45) together in (40) and taking into account conditions on the parameters in Assumptions 5, we conclude that the right hand side of inequality (40) stays finite almost surely, if  $T \rightarrow \infty$  and, thus, almost surely

$$\begin{aligned} 2 \sum_{t=0}^{\infty} \sum_{i=1}^N \beta_i(t+1) (\mathbf{M}_i^0(\boldsymbol{\eta}(t)), \boldsymbol{\mu}_i(t) - \boldsymbol{\mu}_i) \\ + 2 \sum_{t=0}^{\infty} \beta_0(t+1) (-\mathbf{g}(\boldsymbol{\mu}(t)), \boldsymbol{\lambda}(t) - \boldsymbol{\lambda}) < \infty. \end{aligned} \quad (46)$$

Next, we demonstrate that almost surely

$$\lim_{t \rightarrow \infty} (\mathbf{M}^0(\boldsymbol{\eta}(t)), \boldsymbol{\eta}(t) - \boldsymbol{\eta}) \leq 0. \quad (47)$$

Indeed, let us assume that on the contrary, there exists an  $\epsilon > 0$  and  $t_0 > 0$  such that almost surely

$$(\mathbf{M}^0(\boldsymbol{\eta}(t)), \boldsymbol{\eta}(t) - \boldsymbol{\eta}) \geq \epsilon$$

for any  $t \geq t_0$ . In this case, using the idea analogous to the one in (24), we obtain

$$\begin{aligned} 2 \sum_{t=0}^{\infty} \sum_{i=1}^N \beta_i(t+1) (\mathbf{M}_i^0(\boldsymbol{\eta}(t)), \boldsymbol{\mu}_i(t) - \boldsymbol{\mu}_i) \\ + 2 \sum_{t=0}^{\infty} \beta_0(t+1) (-\mathbf{g}(\boldsymbol{\mu}(t)), \boldsymbol{\lambda}(t) - \boldsymbol{\lambda}) \\ \geq \sum_{t=0}^{\infty} 2\beta_{\min}(t+1) (\mathbf{M}^0(\boldsymbol{\eta}(t)), \boldsymbol{\eta}(t) - \boldsymbol{\eta}) \\ - \sum_{t=0}^{\infty} O(\beta_{\max}(t) - \beta_{\min}(t)) \\ \geq \sum_{t=0}^{t_0} \beta_{\min}(t+1) (\mathbf{M}^0(\boldsymbol{\eta}(t)), \boldsymbol{\eta}(t) - \boldsymbol{\eta}) \\ - \sum_{t=0}^{\infty} O(\beta_{\max}(t) - \beta_{\min}(t)) + \epsilon \sum_{t=t_0}^{\infty} \beta_{\min}(t+1) = \infty, \end{aligned}$$

almost surely, according to Assumption 5. This contradicts (46). Thus, (47) holds.

Recall that  $\boldsymbol{\eta}^* \in \mathcal{A} \times \mathbb{R}_+^n$  is a limit point of  $\boldsymbol{\eta}(t)$ , namely  $\lim_{t \rightarrow \infty} \boldsymbol{\eta}(t) = \boldsymbol{\eta}^*$ . Moreover, according to the specification of the constant  $r$  in (38),  $\boldsymbol{\eta}^* \in \mathcal{A} \times \mathbb{R}_{\leq K+r}^n$ . Next, taking into account (47), we obtain:

$$(\mathbf{M}^0(\boldsymbol{\eta}^*), \boldsymbol{\eta} - \boldsymbol{\eta}^*) \geq 0. \quad (48)$$

Since we did not specify the choice of  $\boldsymbol{\eta} \in \mathcal{A} \times \mathbb{R}_{\leq K+r}^n$ , the inequality above holds for any  $\boldsymbol{\eta} \in \mathcal{A} \times \mathbb{R}_{\leq K+r}^n$ . Thus, according to Theorem 6,  $\boldsymbol{\eta}^* = [\boldsymbol{\mu}^*, \boldsymbol{\lambda}^*]$  is a Nash equilibrium in the game  $\Gamma_{ab}(\mathcal{A} \times \mathbb{R}_{\leq K+r}^n)$ . By applying Theorem 5, we conclude that  $\boldsymbol{\mu}^*$  is a generalized Nash equilibrium in the initial game  $\Gamma$ .

Next, we notice that, if  $\boldsymbol{\eta} = \boldsymbol{\eta}^*$  in (39), this inequality (39) together with (41) - (45) imply that

$$\begin{aligned} \mathbb{E}\{\|\boldsymbol{\eta}(t+1) - \boldsymbol{\eta}^*\|^2 | \mathcal{F}_t\} &\leq \|\boldsymbol{\eta}(t) - \boldsymbol{\eta}^*\|^2 \\ &\quad - 2\beta_{\min}(t+1) (\mathbf{M}^0(\boldsymbol{\eta}(t)), \boldsymbol{\eta}(t) - \boldsymbol{\eta}^*) + h(t), \end{aligned} \quad (49)$$

where  $\sum_{t=0}^{\infty} h(t) < \infty$ . Moreover, since  $\mathbf{M}^0$  is pseudo-monotone, (48) implies  $(\mathbf{M}^0(\boldsymbol{\eta}(t)), \boldsymbol{\eta}(t) - \boldsymbol{\eta}^*) \geq 0$  for any  $t$ . Thus, we can apply Theorem 9 to conclude that

$$\|\boldsymbol{\eta}(t) - \boldsymbol{\eta}^*\| \text{ converges almost surely as } t \rightarrow \infty.$$

Since  $\lim_{t \rightarrow \infty} \boldsymbol{\eta}(t) = \boldsymbol{\eta}^* = [\boldsymbol{\mu}^*, \boldsymbol{\lambda}^*]$  almost surely,

$$\lim_{t \rightarrow \infty} \boldsymbol{\eta}(t) = \boldsymbol{\eta}^* = [\boldsymbol{\mu}^*, \boldsymbol{\lambda}^*] \text{ almost surely.}$$

Thus,  $\lim_{t \rightarrow \infty} \boldsymbol{\mu}(t) = \boldsymbol{\mu}^*$ . Finally, Assumption 5 implies that  $\lim_{t \rightarrow \infty} \sigma_i(t) = 0$  for all  $i \in [N]$ . Taking into account that  $\mathbf{x}(t) \sim \mathcal{N}(\boldsymbol{\mu}(t), \boldsymbol{\sigma}(t))$ , we conclude that  $\mathbf{x}(t)$  converges weakly to a Nash equilibrium  $\mathbf{a}^* = \boldsymbol{\mu}^*$  as time runs. Moreover, according to Portmanteau Lemma Klenke (2008), this convergence is also in probability. ■



## V. CONVERGENCE RATE OF THE ALGORITHM

In this section, we estimate convergence rate of the payoff-based procedure presented above in the case of *strongly monotone* (see Definition 2) game mapping  $M$  of the original game  $\Gamma$ . This convergence rate will be expressed in terms of the mean-squared error.

**Theorem 10:** Let the game mapping  $M$  be strongly monotone on  $\mathbf{A}$  with monotonicity constant  $\kappa > 1$ , the function  $\mathbf{g}$  be strongly convex on  $\mathbf{A}$ . Let Assumptions 1-3 hold. Furthermore, assume the time step and variance parameters are chosen according to Remark 4, namely  $\gamma_i(t) = \frac{1}{(t+R_i)^a}$ ,  $\sigma_i(t) = \frac{1}{(t+R_i)^b}$ ,  $i \in [N]$ ,  $\beta_0(t) = \frac{1}{(t+N_0)^{a+2b}}$ , where  $a + 2b \in (0.5, 1]$ ,  $2a > 1$ , and  $a + 3b > 1$ . Then in the long run of the algorithm (11)-(12)

$$\mathbb{E}\{\|\mu(t) - \mu^*\|^2\} \leq \frac{C}{t^{2(a+b)-1}} = O(1/t^{2(a+b)-1}),$$

where  $C$  is some positive constant and  $\mu^*$  is the unique variational equilibrium of the game  $\Gamma$  to which the vector  $\mu(t)$  converges almost surely.

First, we prove a lemma that will be used in the proof of the theorem above.

**Lemma 3:** Let the sequence  $\{a_t\}$ ,  $a_t \geq 0$   $t \in \mathbb{Z}_+$ , satisfy the following iteration:

$$a_{t+1} \leq (1 - \kappa/t)a_t + \psi/t^c,$$

for some constants  $1 < c \leq 2$ ,  $\kappa > 1$ ,  $\psi > 0$ . Then

$$a_t \leq \frac{C}{t^{c-1}},$$

where  $C = \max\{a_0, \frac{\psi}{\kappa-1}\}$ .

**Proof:** The proof is based on a standard rate estimate, analogous to the result in (5.292) in Shapiro *et al.* (2014). We prove the claim by induction. Let us assume that  $a_0 \geq \frac{\psi}{\kappa-1}$ . Then, according to the induction step,

$$a_{t+1} \leq \left(1 - \frac{\kappa}{t}\right) \frac{a_0}{t^{c-1}} + \frac{\psi}{t^c}.$$

Thus, it suffices to show that the right hand side of the inequality above is not more than  $\frac{a_0}{(t+1)^{c-1}}$ . As  $\frac{\psi}{a_0} - \kappa \leq -1$ ,

$$\begin{aligned} \left(1 - \frac{\kappa}{t}\right) \frac{a_0}{t^{c-1}} + \frac{\psi}{t^c} &= \frac{a_0}{t^c} \left(t - \kappa + \frac{\psi}{a_0}\right) \leq \frac{a_0}{t^c} (t-1) \\ &\leq \frac{a_0}{(t+1)^{c-1}}, \end{aligned}$$

since  $a_0 > 0$  and

$$\left(1 + \frac{1}{t}\right)^{c-1} \leq \left(1 + \frac{1}{t-1}\right).$$

The case  $a_0 \leq \frac{\psi}{\kappa-1}$  can be considered analogously.  $\blacksquare$

**Proof: (of Theorem 10)** First, let us notice that, according to Theorem 4, the Nash equilibrium  $\eta^* = [\mu^*, \lambda^*]$  of the associated bounded game  $\Gamma_{ab}(\mathbf{A} \times \mathbb{R}_{\leq K+r}^n)$  is the solution of  $VI(\mathbf{A} \times \mathbb{R}_{\leq K+r}^n, M^0)$ . Hence, according to the main property of solutions of variational inequalities Pang & Facchinei (2003), for any  $i \in [N]$

$$\mu_i^* = \text{Proj}_{A_i}[\mu_i^* - \beta_i(t)M_i^0(\eta^*)],$$

Next, we estimate  $\|\mu(t+1) - \mu^*\|^2$ . Taking into account the equality above, (17), and non-expansion property of the projection operator, we obtain that there exists  $t_0$  such that for any  $t \geq t_0$  and some  $\psi > 0$

$$\begin{aligned} \mathbb{E}\|\mu(t+1) - \mu^*\|^2 &\leq \mathbb{E}\|\mu(t) - \mu^*\|^2 \\ &\quad - 2\beta_{\min}(t)\mathbb{E}(M_{(N)}^0(\eta(t)) - M_{(N)}^0(\eta^*), \mu - \mu^*) \\ &\quad + O(\sigma_{\max}^2(t)\gamma_{\max}^2(t)) \\ &\leq \left(1 - \frac{\kappa}{t}\right) \mathbb{E}\|\mu(t) - \mu^*\|^2 + \frac{\psi}{t^{2(a+b)}}, \end{aligned} \quad (50)$$

where

$$M_{(N)}^0(\eta(t)) = [M_1^0(\eta(t)), \dots, M_N^0(\eta(t))].$$

To get the first inequality in (50), we used boundedness of  $\eta(t)$  (see Lemma 2), an estimation analogous to the one in (39), the fact that  $\mathbb{E}\{R_i(\mathbf{x}(t), \eta(t))\} = \mathbb{E}\{E\{R_i(\mathbf{x}(t), \eta(t))|\mathcal{F}_t\}\} = 0$ , and properties of terms in (41) - (45). For the second inequality, we used the strong monotonicity of the map  $M$  over  $\mathbf{A}$ , which implies strong monotonicity of  $M_{(N)}^0$  over  $\mathbf{A}$ , since  $\mathbf{g}$  is convex on  $\mathbf{A}$  and  $\lambda(t) \in \mathbb{R}_+^n$  (see eq. (3)). Moreover, according to definition of  $\beta_i$ , there exists  $t_0$  such that  $\beta_{\min}(t) \geq \frac{1}{2t}$  for any  $t \geq t_0$ . Finally, taking into account that  $1 < 2(a+b) \leq 2$  and using Lemma 3 for  $c = 2(a+b)$ , we conclude that  $\mathbb{E}\{\|\mu(t) - \mu^*\|^2\} \leq C/t^{2(a+b)-1}$ , where  $C = \max\{\mathbb{E}\{\|\mu(0) - \mu^*\|^2\}, \frac{\psi}{\kappa-1}\}$ .  $\blacksquare$

The result above demonstrates that the convergence rate of the proposed payoff-based learning procedure is sublinear, which is consistent with results on the related optimization algorithms and procedures for Nash equilibrium learning based on the stochastic approximation technique Juditsky *et al.* (2011); Tatarenko & Touri (2017). Note also that Theorem 10 presents the asymptotic estimation of the convergence rate, whereas its tightness and more details on characterization of the constant  $C$  need to be analyzed separately and are subject of our future work.

## VI. CONCLUSION

We proposed a novel decentralized payoff-based learning approach to compute the variational Nash equilibria in multi-agent convex games with convex coupling constraints. In this approach, each agent determined its next state by sampling from a Gaussian distribution, whose mean was considered to be the agent's local action and was updated using the payoff information. We proved almost sure convergence of the means of the distributions, and, hence, of the agents' joint actions, to a generalized Nash equilibrium, given appropriate choice of the sequence of the step sizes and the variances of the distributions. Furthermore, we quantified the convergence rate of this payoff-based algorithm.

Future work can explore existence of other payoff-based algorithms with potentially faster convergence rates. Furthermore, it would be interesting to generalize the results to non-convex games on a network, where agents can communicate with each other according to a graph representing a communication topology. Finally, an interesting relevant future work can develop payoff-based algorithms for the case of non-jointly convex coupling constraints.

## REFERENCES

- Arslan, G., Marden, J. R., & Shamma, J. S. 2007. Autonomous vehicle-target assignment: a game theoretical formulation. *ASME Journal of Dynamic Systems, Measurement and Control*, **129**(September), 584–596.
- Bharath, B., & Borkar, V. S. 1999. Stochastic approximation algorithms: Overview and recent trends. *Sadhana*, **24**(4), 425–452.
- Boyd, S., & Vandenberghe, L. 2004. *Convex optimization*. Cambridge university press.
- Dutta, P. S., Jennings, N. R., & Moreau, L. 2011. Cooperative Information Sharing to Improve Distributed Learning in Multi-Agent Systems. *CoRR*, **abs/1109.5712**.
- Facchinei, F., & Kanzow, C. 2007. Generalized Nash equilibrium problems. *4OR*, **5**(3), 173–210.
- Facchinei, F., Fischer, A., & Piccialli, V. 2007. On generalized Nash games and variational inequalities. *Operations Research Letters*, **35**(2), 159 – 164.
- Goto, T., Hatanaka, T., & Fujita, M. 2012 (June). Payoff-based Inhomogeneous Partially Irrational Play for potential game theoretic cooperative control: Convergence analysis. *Pages 2380–2387 of: American Control Conference (ACC), 2012*.
- Gowda, M. S. 1990. Affine Pseudomonotone Mappings and the Linear Complementarity Problem. *SIAM Journal on Matrix Analysis and Applications*, **11**(3), 373–380.
- Jensen, M. K. 2010. Aggregative Games and Best-Reply Potentials. *Economic Theory*, **43**(1), 45–66.
- Juditsky, A., Nemirovski, A., & Tauvel, C. 2011. Solving variational inequalities with stochastic mirror-prox algorithm. *Stoch. Syst.*, **1**(1), 17–58.
- Kannan, A., & Shanbhag, U. V. 2012. Distributed Computation of Equilibria in Monotone Nash Games via Iterative Regularization Techniques. *SIAM Journal on Optimization*, **22**(4), 1177–1205.
- Klenke, A. 2008. *Probability theory: a comprehensive course*. London: Springer.
- Li, N., & Marden, J. R. 2013. Designing Games for Distributed Optimization. *IEEE Journal of Selected Topics in Signal Processing*, **7**(2), 230–242. Special issue on adaptation and learning over complex networks.
- Marden, J. R., & Shamma, J. S. 2012. Revisiting log-linear learning: Asynchrony, completeness and payoff-based implementation. *Games and Economic Behavior*, **75**(2), 788 – 808.
- Marden, J. R., Arslan, G., & Shamma, J. S. 2009a. Cooperative control and potential games. *Trans. Sys. Man Cyber. Part B*, **39**(6), 1393–1407.
- Marden, J. R., Young, H. P., Arslan, G., & Shamma, J. S. 2009b. Payoff-Based Dynamics for Multiplayer Weakly Acyclic Games. *SIAM J. Control and Optimization*, **48**(1), 373–396.
- Marden, J. R., Ruben, S. D., & Pao, L. Y. 2013. A Model-Free Approach to Wind Farm Control Using Game Theoretic Methods. *IEEE Trans. Contr. Sys. Techn.*, **21**(4), 1207–1214.
- Nevelson, M. B., & Khasminskii, R. Z. 1973. *Stochastic approximation and recursive estimation [translated from the Russian by Israel Program for Scientific Translations ; translation edited by B. Silver]*. American Mathematical Society.
- Paccagnan, D., Gentile, B., Parise, F., Kamgarpour, M., & Lygeros, J. 2016a (Dec). Distributed computation of generalized Nash equilibria in quadratic aggregative games with affine coupling constraints. *Pages 6123–6128 of: IEEE Conference on Decision and Control*.
- Paccagnan, D., Kamgarpour, M., & Lygeros, J. 2016b (June). On aggregative and mean field games with applications to electricity markets. *Pages 196–201 of: 2016 European Control Conference (ECC)*.
- Pang, J.-S., & Facchinei, F. 2003. *Finite-dimensional variational inequalities and complementarity problems : vol. 1*. Springer series in operations research. New York, Berlin, Heidelberg: Springer.
- Perkins, S., Mertikopoulos, P., & Leslie, D. S. 2015. Mixed-Strategy Learning with Continuous Action Sets. *IEEE Transactions on Automatic Control*.
- Poljak, B. T. 1987. *Introduction to optimization*. Optimization Software.
- Pradelski, B., & Young, H. P. 2012. Learning efficient Nash equilibria in distributed systems. *Games and Economic behavior*, **75**(2), 882–897.
- Rosen, J. B. 1965. Existence and Uniqueness of Equilibrium Points for Concave N-Person Games. *Econometrica*, **33**(3), 520–534.
- Saad, W., Zhu, H., Poor, H. V., & Basar, T. 2012. Game-theoretic methods for the smart grid: An overview of microgrid systems, demand-side management, and smart grid communications. *IEEE Signal Processing Magazine*, **29**(5), 86–105.
- Salehisadaghiani, F., & Pavel, L. 2014 (Dec). Nash equilibrium seeking by a gossip-based algorithm. *Pages 1155–1160 of: 53rd IEEE Conference on Decision and Control*.
- Scutari, G., Barbarossa, S., & Palomar, D. P. 2006 (May). Potential Games: A Framework for Vector Power Control Problems With Coupled Constraints. *Pages 241–244 of: 2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, vol. 4.
- Scutari, G., Palomar, D. P., Facchinei, F., & Pang, J.-S. 2012. Monotone games for cognitive radio systems. *Pages 83–112 of: Distributed Decision Making and Control*. Springer.
- Shapiro, A., Dentcheva, D., & Ruszczyński, A. 2014. *Lectures on Stochastic Programming: Modeling and Theory, Second Edition*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics.
- Tatarenko, T. 2014 (June). Proving convergence of log-linear learning in potential games. *Pages 972–977 of: American Control Conference (ACC), 2014*.
- Tatarenko, T. 2016a (Dec). Stochastic payoff-based learning in multi-agent systems modeled by means of potential games. *Pages 5298–5303 of: 2016 IEEE 55th Conference on Decision and Control (CDC)*.
- Tatarenko, T. 2016b (June). Stochastic stability of potential function maximizers in continuous version of independent log-linear learning. *Pages 210–215 of: 2016 European Control Conference (ECC)*.
- Tatarenko, T., & Kamgarpour, M. 2017. Payoff-Based Ap-

- proach to Learning Nash Equilibria in Convex Games. *In: The 20th World Congress of the International Federation of Automatic Control, IFAC*. submitted.
- Tatarenko, T., & Touri, B. 2017. Non-Convex Distributed Optimization. *IEEE Transactions on Automatic Control*, **PP**(99), 1–1.
- Tellidou, A. C., & Bakirtzis, A. G. 2007. Agent-Based Analysis of Capacity Withholding and Tacit Collusion in Electricity Markets. *IEEE Transactions on Power Systems*, **22**(4), 1735–1742.
- Thathachar, A. L., & Sastry, P. S. 2003. *Networks of Learning Automata: Techniques for Online Stochastic Optimization*. Springer US.
- Yin, H., Shanbhag, U. V., & Mehta, P. G. 2011. Nash Equilibrium Problems With Scaled Congestion Costs and Shared Constraints. *IEEE Transactions on Automatic Control*, **56**(7), 1702–1708.
- Zhu, M., & Frazzoli, E. 2016. Distributed robust adaptive equilibrium computation for generalized convex games. *Automatica*, **63**, 82 – 91.
- Zhu, M., & Martínez, S. 2013. Distributed Coverage Games for Energy-Aware Mobile Sensor Networks. *SIAM J. Control and Optimization*, **51**(1), 1–27.